

博士論文

ソフトウェア開発の超上流工程における
非機能要件の定量的評価

齊藤康廣

2015年2月3日

奈良先端科学技術大学院大学
情報科学研究科 情報システム学専攻

本論文は奈良先端科学技術大学院大学情報科学研究科に
博士(工学) 授与の要件として提出した学位論文である。

齊藤康廣

審査委員：

松本健一 教授 (主指導教員)

藤川和利 教授 (副指導教員)

門田暁人 准教授 (副指導教員)

ソフトウェア開発の超上流工程における 非機能要件の定量的評価*

齊藤康廣

内容梗概

委託ソフトウェア開発プロジェクトの超上流工程において、ユーザ(発注企業)が作成する提案依頼書(Request for Proposal: RFP)の品質は、ソフトウェア開発を成功させる上で極めて重要である。RFPは多様な情報で構成されているが、品質評価の重要な対象の一つとなるのが、非機能要件(Non Functional Requirements: NFR)である。NFRは、開発すべきソフトウェアのアーキテクチャや、保守・運用における制約条件となるため、NFRが明確に記述されているか否かは、RFP品質を議論する上で特に重要な観点となる。

本論文では、RFPの品質を定量的に評価する3つの方法を提案する。評価対象とするのは、RFPで示されるべき非機能要件(NFR)であり、評価の観点は、その記述の明確さである。

最初の提案では、評価対象を、ユーザにとって重要度の高い「保守と運用に関する55個の非機能要件」に限定した上で、要件記述の明確さを最大5段階で評価するためのメトリクスを定義する。評価結果は、RFPの総合評価点と、要件毎の評価点を俯瞰するためのレーダチャートとして示される。地方自治体、図書館、官公庁、独立行政法人、大学、病院の6ドメイン29件のRFPを評価対象としたケーススタディによって、記述が不十分な要件を特定することにより、基準値との比較を通じて特に改善が必要な特性を明らかにできることを確認した。

2番目の提案では、評価の自動化を目的として、RFPに含まれる各NFRに関するキーワード(NFRキーワード)の出現頻度に基づいて、教師あり機械学習によって各要件の記述の

* 奈良先端科学技術大学院大学 情報科学研究科博士論文, NAIST-IS-DD1061202, 2015年2月3日.

明確さを評価する。提案方法では、自然言語で記述されたRFPからNFRキーワードを抽出し、各NFRにマッピングする。そして、NFRキーワードの出現頻度とNFRの記述の明確さとの関係をランダムフォレストによりモデル化する。70件のRFPを題材として、提案方法によって26種類の非機能要件の記述の明確さを評価した結果、人手による評価との一致率の平均値は69.8%であり、±1の誤差を許容した±1差一致率の平均値は97.2%となった。

3番目の提案では、評価の自動化において教師データを不要とすることを目的として、NFRキーワードの出現頻度とそれに対する重み付けのみに基づいて、関連するNFRに評価値を与える。提案方法では、重み付けを与える方法として、単純TF-IDF、拡張TF-IDF、およびトークン比の3種類を採用し、実験による比較を行った。その結果、拡張TF-IDFがもっともよい結果を示し、提案方法による評価値と人手による評価値との相関係数は0.22～0.43となった。このことから、教師データを付与することが難しい状況においても、ある程度の精度でNFRの自動評価を行えることが示唆された。

キーワード

提案依頼書, 非機能要件, 評価指標, 機械学習, テキストマイニング

Quantitative Evaluation of Non Functional Requirements in an Early Stage of Software Development*

Yasuhiro Saito

Abstract

In the beginning of a contracted based software development project, the quality of a Request For Proposal (RFP) is extremely important for successful development of a software system. The RFP is a document provided by a software user company and used as an initial system requirements specification to ask software developer companies to propose their technical plans to fulfill the requirements. The RFP consists of various requirements information, and one of the most important information is Non Functional Requirements (NFRs). It is because NFRs define constraints of software architecture, and they also affect the software maintenance cost and operation cost. Therefore, it is extremely important to evaluate whether NFRs are written clearly enough or not.

In this thesis, three methods for quantitative evaluation of an RFP are proposed. The goal of evaluation is clarity of description of NFRs. Firstly, we selected a set of fifty-five important NFRs that should be written in an RFP. Then, metrics to evaluate the clarity of NFRs was defined. The result of evaluations are shown by total score and radar charts for the purpose to observe low clarity NFRs. In a case study, RFPs from 29 projects were evaluated by the proposed metrics. As a result, we confirmed that the identification of poorly-written NFRs became easily recognizable by the proposed method.

*Doctoral Dissertation, Department of Information Systems, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-DD1061202, February 3, 2015.

Secondly, for the purpose of automatic evaluation of clarity of NFR descriptions, we proposed a machine learning approach using Random Forest and term frequency of NFR key words extracted by text mining. An estimation model trained from known NFRs automatically gives the clarity of NFRs in the unknown RFPs. As a result of an evaluation of 26 NFR categories using 70 RFPs, 69.8% of the estimation results showed identical score with human evaluation. And, 97.2% of the estimation results matched human score if we allow plus or minus one score difference.

Thirdly, for the purpose of automatic evaluation of without training data, we proposed an unsupervised technique to evaluate the clarity of NFR descriptions. The proposed method estimates the clarity based on term frequency of key words related to NFRs. Three types of weights to key words are used, that is, TF-IDF, extended TF-IDF and key words ratio. As a result of evaluation of the proposed method and comparison of weight types using one hundred sixty one RFPs, extended TF-IDF showed the best fit to human evaluation, and the correlation coefficient between estimated clarity and human evaluation ranged from 0.22 to 0.43. The result shows that the automatic evaluation is possible to some extent even if one cannot prepare training data for supervised machine learning.

Keywords:

Request For Proposal, Non-Functional Requirements, Machine Learning, Text Mining

関連発表論文

学術論文誌

1. 齊藤康廣, 門田暁人, 松本健一. “非機能要件に着目した Request For Proposal (RFP) 評価”. SEC Journal, 第 38 号, September 2014, pp. 30-37. (第 2 章に関連する)

国際会議発表

1. Yasuhiro Saito, Akito Monden, Kenichi Matumoto. “Evaluation of Non Functional Requirements in a Request for Proposal (RFP)” . Proceedings of 2012 Joint Conf of 22nd Int'l Workshop on Software Measurement and 7th Int'l Conference on Software Process and Product Measurement (IWSM-MENSURA), October 2012, pp.106-111. (第2章に関連する)

国内研究集会発表

1. 齊藤康廣, 門田暁人, 松本健一. “Request For Proposal (RFP) における保守・運用要件指標の抽出と評価”. 情報処理学会研究報告 2012-SE-175(1), March 2012, pp.1-10. (第2章に関連する)
2. 齊藤康廣, 門田暁人, 松本健一. “RFP における機械学習による非機能要件の評価”. 情報処理学会研究報告. 2013-SE-179(5), March 2013, pp.1-7. (第 3 章に関連する)

目次

第1章 序論	1
1.1 研究の背景	1
1.2 研究の目的	2
1.3 論文構成	4
第2章 非機能要件に着目したRFPの評価	6
2.1 概要	6
2.2 関連研究	7
2.2.1 非機能要件に関するガイドライン	7
2.2.2 RFPや非機能要件に関する従来研究	8
2.3 提案方法	8
2.3.1 概要	8
2.3.2 評価対象とする非機能要件	9
2.3.3 非機能要件評価シート	9
2.3.4 評価結果	13
2.4 ケーススタディ	13
2.4.1 概要	13
2.4.2 総合評価点	14
2.4.3 レーダチャート	15
2.4.4 ベンチマーキング	17
2.4.5 評価者間の評価点のばらつき	19
2.5 まとめ	21
2.5.1 評価結果と考察	21
2.5.2 課題	22
第3章 機械学習による非機能要件の自動評価	24
3.1 概要	24
3.2 関連研究	24

3.3	非機能要件キーワードの出現頻度予備分析	26
3.3.1	NFR キーワードの抽出	26
3.3.2	「小項目」と NFR キーワードの関連付け	28
3.3.3	「小項目」と NFR キーワード出現頻度	30
3.4	提案手法	33
3.4.1	「小項目」の自動評価モデル構築の概要	33
3.4.2	非機能要件評価モデルの構築	35
3.4.3	「NFR キーワード評価シート」の利用	39
3.5	ランダムフォレストを用いた非機能要件評価実験	39
3.5.1	実験の目的	39
3.5.2	実験データと実験方法	40
3.5.3	実験結果の評価手法.	42
3.5.4	実験結果に基づく一致の検定と考察.	44
3.5.5	手動評価と自動評価の比較.	46
3.6	まとめ	51
第 4 章	機械学習によらない非機能要件の自動評価	53
4.1	概要	53
4.2	関連研究	53
4.3	提案手法	55
4.3.1	NFR キーワードの抽出	55
4.3.2	NFR キーワードの重み	55
4.3.3	単純及び拡張 TF-IDF とキーワード比	56
4.3.4	RFP の NFR 評価スコア	60
4.4	評価実験	62
4.4.1	評価実験データと「小項目」	63
4.4.2	評価実験と考察	63
4.5	クラスタリングによる評価スコアの段階評価.	74

4.5.1	提案手法	74
4.5.2	クラスタリング評価実験	76
4.5.3	評価実験結果と考察	76
4.6	まとめ	81
第5章	社会的影響別の非機能要件評価	83
5.1	概要	83
5.2	社会的影響による RFP の分類と手動評価	83
5.3	社会的影響カテゴリ別の評価スコア	86
5.4	社会的影響カテゴリ別の評価比較	89
5.5	まとめ	91
第6章	結論	92
	謝辞	
	参考文献	

目 次

図1.1	委託開発ソフトウェア開発プロジェクト	2
図2.1	提案するRFP評価法の概要	12
図2.2(a)	5段階評価 明確さ評価基準の例	12
図2.2(b)	3段階評価 明確さ評価基準の例	12
図2.2(c)	2段階評価 明確さ評価基準の例	12
図2.3(a)	総合評価点 システムドメイン毎の評価点	16
図2.3(b)	レーダチャート:大項目 システムドメイン毎の評価点	16
図2.3(c)	レーダチャート:中項目 システムドメイン毎の評価点	17
図2.4	ケーススタディ結果:基準値(RFPトップ3)との比較	19
図3.1	NFRキーワード評価シート(抜粋)	28
図3.2	「小項目」ごとのNFRキーワード出現頻度	29
図3.3	RFP語数とNFRキーワード出現頻度	31
図3.4	NFRキーワード出現頻度とRFPの手動評価ポイント	32
図3.5	NFRキーワード出現頻度と「小項目」の種類	33
図3.6	機械学習によるNFRキーワード評価モデル構築手順	35
図3.7	「小項目」の3段階評価の比較	45
図3.8	クロス表と κ 統計量(一部)	46
図3.9(a)	自動評価と手動評価の評価得点レーダチャート(3段階評価)	50
図3.9(b)	自動評価と手動評価の評価得点レーダチャート(5段階評価)	51
図4.1	単純TF-IDFとNFRキーワードの出現頻度	59
図4.2	拡張TF-IDFとNFRキーワード出現RFP数	60
図4.3(a)	応答性のNFRキーワード評価別箱ひげ図	66
図4.3(b)	負荷バランスのNFRキーワード評価別箱ひげ図	66
図4.3(c)	リカバリ処理のNFRキーワード評価別箱ひげ図	67
図4.3(d)	アクセス管理のNFRキーワード評価別箱ひげ図	67
図4.3(e)	セキュリティ管理のNFRキーワード評価別箱ひげ図	68
図4.3(f)	障害対策のNFRキーワード評価別箱ひげ図	68

図4.4(a)	手動評価点毎の拡張TF-IDF評価点の比率:応答性.	70
図4.4(b)	手動評価点毎の拡張TF-IDF評価点の比率:負荷バランス.	70
図4.4(c)	手動評価点毎の拡張TF-IDF評価点の比率:リカバリ処理.	71
図4.4(d)	手動評価点毎の拡張TF-IDF評価点の比率:アクセス管理.	71
図4.4(e)	手動評価点毎の拡張TF-IDF評価点の比率:セキュリティ管理.	72
図4.4(f)	手動評価点毎の拡張TF-IDF評価点の比率:障害対策	72
図4.5	手動評価別のNFRキーワード出現回数と拡張TF-IDF評価点	73
図4.6	手動3段階評価別の拡張TF-IDF評価スコア.	73
図4.7	「小項目」ごとのNFR評価スコア一致数.	78
図4.8	「小項目」別の一致率および不一致率.	81
図5.1	社会的影響カテゴリ別の「小項目」の手動評価スコア.	85
図5.2	社会的影響カテゴリ別の個別RFPの小項目評価スコア.	87
図5.3	社会的影響カテゴリ別の個別RFP評価	89
図5.4(a)	社会的影響カテゴリ別の小項目手動評価スコア	90
図5.4(b)	社会的影響カテゴリ別の小項目拡張TF-IDF評価スコア.	90

表 目 次

表2.1	各評価者の各RFPに対する総合評価点.	21
表 3.1	NFR キーワードの重要度と出現頻度.	27
表 3.2	出現した RFP 数が多い上位 20 個の NFR キーワード.	30
表 3.3	NFR キーワードと RFP の相関関係	33
表 3.4	段階評価とポイント	36
表 3.5	NFR キーワード評価シート(抜粋)	38
表 3.6	段階評価法別の κ 統計量	46
表 3.7(a)	ランダムフォレストの評価推定(3 段階評価)	48
表 3.7(b)	ランダムフォレストの評価推定(5 段階評価).	49
表 3.8	自動評価と手動評価の誤差.	50
表 4.1	3つの手法による重みと NFR キーワードの出現頻度(抜粋).	59
表 4.2	NFR キーワード重みの例.	62
表 4.3	NFR キーワード重みと手動評価の相関係数	64
表 4.4	NFR キーワードに関連する記述例「障害対策」.	65
表 5.1	社会的影響カテゴリ別の手動評価(一部の NFR 評価指標)	85

付録 目 次

付録 A.	「NFR キーワード評価シート」	102
付録 B.	161 件の RFP による「NFR キーワード評価シート」の評価対象「小項目」	108

第 1 章 序論

1.1 研究の背景

提案依頼書(Request For Proposal, 以後は RFP とする)は, ソフトウェア開発を委託するにあたり, 委託元企業(ユーザ)が, 委託先候補の企業(ベンダ)に対して, 開発に関する具体的な技術提案(技術仕様・技術提案書の作成)を依頼する文書である. RFP には, 機能要件, 非機能要件, 事務要件, システム要件, ライセンス事項, 開発者資格, 契約要件などが記述されている. ユーザは, 提示された技術仕様・技術提案書に基づいてベンダを選定し, 契約仕様書の作成, 契約の締結を経て, ソフトウェアの開発作業が開始されることになる. RFP は, ソフトウェアの委託開発のベースとなる, 重要な文書の一つであり, その品質が, ソフトウェア開発の成否を大きく左右することになる[8] .

委託ソフトウェア開発の超上流工程では, ユーザが作成した RFP に記述されている要求仕様をもとにベンダが技術提案書を作成し, 両者が技術協議を通じてユーザ要求を明確にしていく. このプロセス(図 1.1 参照)は委託ソフトウェアの開発においてユーザ要求を後工程のソフトウェア設計において明確に反映するために行われる. しかし, 多くの委託ソフトウェア開発の超上流工程において, ユーザとベンダ間で要件が十分に確認されずに契約が行われ, ソフトウェア開発以後の工程でユーザ要件に係る問題が発生する. 委託ソフトウェア開発では, 要件定義が不十分であると, 「スコープ・クリープ(要求のなしくずし的な追加・変更)」が問題となり, ソフトウェア品質や納期に影響を与える[50] . 特に非機能要件については, 後工程での追加・変更が困難であるが, 文献によると「非機能要求については, ユーザは要求仕様にはほとんど何も記述していない. これが原因になって, 後になってユーザとベンダ間の紛争の種になる」[3]と言われている. ここで「非機能要件」とは, 「非機能要求は, 解に対して得られる制約のことである. 非機能要求は, しばしば制約または品質要求という名で知られている. 「SWEBOK2004」[3]と定義されている非機能要求を要件として記述されたものとする. 「非機能要件」は, ユーザとベンダ双方にとって, ソフトウェア品質の物理的な制約条件を合意することであるから, 「機能要件」と比較してより

完全性と網羅性を契約前に明確にしておく必要がある。よって、この問題について、ユーザとベンダ間の技術協議にて、ユーザ要件の客観的な評価を通じて議論を行うことが重要である。従来研究では、ソフトウェア開発プロセスを通じて定量的な手法を導入することにより、ソフトウェア開発プロジェクトをソフトウェアメトリクスの観点から指標を計測し、管理することが行われている[46][47]。しかし、多くの場合、実際のソフトウェア設計及び製作の段階以後を対象としており、超上流工程での要件定義におけるメトリクスに関する関連研究は極めて少ない[27][48]。ソフトウェアライフサイクルの最も初期段階である超上流工程の要件定義にソフトウェアメトリクスを導入することによりユーザ要求を計測することは、ソフトウェア開発プロセスを管理するためにも極めて重要であると考えられる。

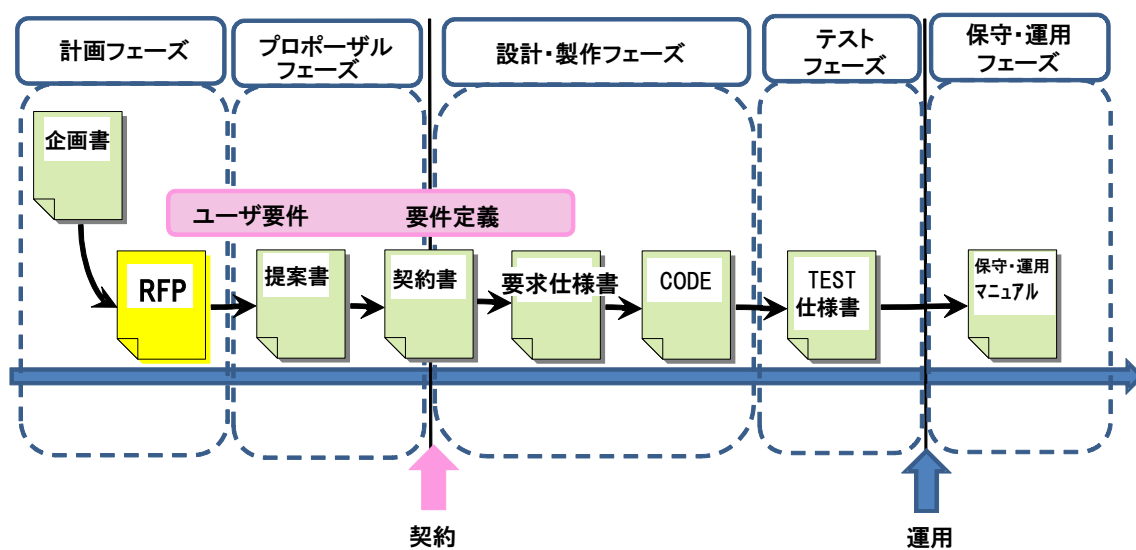


図 1.1 委託開発ソフトウェア開発プロジェクト

1.2 研究の目的

本論文は、ユーザが作成した RFP に記述されている非機能要件 (以後適宜 NFR とする) の記述内容の明確さを定量評価する評価指標を提案し、実際の RFP をケーススタディとして評価実験を行うことによりその有効性を評価することを目的とする。

ソフトウェアライフサイクルの超上流工程で最も重要な書類の一つである RFP に記述されているユーザの非機能要求を評価するためには、RFP に記述されるべき「非機能要件」について明らかにする必要がある。「非機能要件」を定義するには、まず、第一にユーザ要求の把握が必要である。ただし、ユーザが RFP を作成した段階で、すべての非機能要求を把握しているとは限らず、ユーザが記載している内容は、ユーザのドメイン知識に基づいた達成すべきシステムへの要求であり、そのソフトウェアシステムを実現するための要件とは限らない場合がある。これらに対して、ベンダはソフトウェアシステムを構築する上で求められる「非機能要件」を提案する必要がある。非機能要求として、共通的に重視される特性として、「性能(効率性)」、「信頼性」、「セキュリティ」、「使用性」および「保守性」が指摘されている[49]。しかし、これらの特性は「非機能要件」としての大きな特性であり、また、さまざまな語句が使用されている[3]。さらに、RFP に記述された「非機能要件」を定量評価する上で必ずしも適当な特性であるとは言えない。よって、文献や従来研究の調査により、RFP に記述されるべき「非機能要件」の記述内容の明確さを評価する特性を選別する必要がある。

RFP にはユーザの機能要求および非機能要求が自然言語で記述されており、ユーザ要件の記述があいまいであることが要因となり後工程のソフトウェア設計において要求の実現されたソフトウェアが構築できないことが問題となる。そこで、この問題解決を支援するモデルとして、最初に、文献や市場調査の結果を参考にした RFP の非機能要件指標(以後 NFR 指標とする)で構成される「非機能要件評価シート」を提案する。2 番目に、「非機能要件」の自動評価を目的として、自然言語で記述された RFP から NFR 指標に関連する NFR キーワードを抽出して「NFR キーワード評価シート」を作成する。そして、NFR キーワードの出現頻度にもとづいて算出した評価点による自動評価モデルを提案する。自動評価モデルでは、教師あり機械学習によるランダムフォレストによる 70 件の RFP について評価推定の実験を行う。自動評価モデルによる実験結果は、3つの段階評価結果を人手による評価(以後手動評価とする)と比較し、自動評価に有効な段階評価について考察する。3 番目に、多くの RFP を題材として評価の自動化を行う場合には、手動評価による教師データの作成に時間

を要することから、教師なしで自動評価する 3 つの重み付け手法の有効性を評価する。TF-IDF 手法、TF-IDF を拡張した拡張 TF-IDF およびトークン比の考えによる 3 つの手法で NFR キーワードに重み付けした評価点による比較を行う。評価実験により、3 つの手法を手動評価の結果と比較し、最も有効な手法を提案する。さらに、教師なし機械学習による自動評価を目的として、評価スコアをクラスタリング手法により 3 グループに分類し、手動評価による 3 段階評価との比較を行う。最後に、「NFR キーワード評価シート」による評価が情報化対象ドメインによらず有効であることを確認するために、情報化対象システムをその社会的影響により 3 つのカテゴリに分類した 161 件の RFP について評価点を算出し、比較評価を行う。

提案した評価モデルによる定量評価を行い、評価結果に基づいて技術協議を進めることで、ソフトウェア開発の超上流工程でのユーザとベンダの合意形成が促進され、RFP の記述内容の明確さを向上することが可能であると考えられる。また、ユーザ要求の変化を定量評価結果の時系列データの推移として管理することにより、さらに後工程であるソフトウェア開発工程においても、ユーザ要件について、ユーザとベンダ間の合意を支援する有効なツールとなり得ると考える。

1.3 論文構成

本論文は大きく 4 つの章から構成される。第 2 章では、RFP の非機能要件を評価する「非機能要件評価シート」を提案し、29 件の RFP についてケーススタディを行う。「非機能要件評価シート」にもとづいて手動評価した定量評価が、非機能要件の記述の明確さについて有効であるかどうかを確認する。第 3 章では、自然言語で記述された 70 件の RFP から非機能要件を記述した NFR キーワードを抽出する。抽出した NFR キーワードを第 2 章で作成した「非機能要件評価シート」と対応させた「NFR キーワード評価シート」を提案する。「NFR キーワード評価シート」の NFR キーワード出現頻度にもとづいて非機能要件の記述内容の明確さを、教師あり機械学習のランダムフォレストにより推定する評価実験を行う。ランダムフォレストによる評価推定と手動評価の一致度合いを評価

し、自動評価の有効性を検証する。第 4 章では、教師なし自動評価の有効性を確認するために、NFR キーワードの重み付けとして NFR キーワード出現頻度にもとづく「単純 TF-IDF」「拡張 TF-IDF」および「キーワード比」の 3 つの手法による評価実験を行う。手動評価との比較により、3 つの手法の中から有効な手法を提案する。さらに、追加実験として、教師なし機械学習によるクラスタリング手法の一つである k-means 法により自動評価された結果と 3 段階の手動評価との評価実験を行い、教師なし機械学習による評価の有効性を検証する。第 5 章では、161 件の RFP を社会的影響で 3 つに分類した情報化対象システムについて、「NFR キーワード評価シート」を適用することにより、非機能要件の記述内容の明確さが有効に評価されることを確認する。最後に、第 6 章で論文全体の結論を述べ、まとめを行う。

第 2 章 非機能要件に着目した RFP の評価

2.1 概要

RFP は多様な情報で構成されているが、品質評価の重要な対象の一つとなるのが、「非機能要件 (Non Functional Requirements: NFR)」である。NFR は、開発すべきソフトウェアのアーキテクチャに対する制約条件となり、アーキテクチャの実現可能性に大きく影響する。アーキテクチャは、ソフトウェア品質を決定する主要因の一つとされている[5]。更に、開発開始後のアーキテクチャ変更が容易でないことから、RFP に基づく技術仕様・技術提案書の作成において、アーキテクチャの策定やその実現可能性の評価は、ベンダにとって極めて重要な作業の一つとなっている。NFR が明確に記述されているか否かは、RFP 品質を議論する上で重要な観点の一つと言える。

本章では、ベンダへの提案依頼書 (RFP) 提示に先立ち、RFP 作成者であるユーザ自身が、RFP の品質を定量的に評価する方法を提案する。評価対象とするのは、RFP で示されるべき非機能要件 (NFR) であり、評価の観点は、その記述の明確さ、である。RFP に記述すべき NFR を示すガイドラインや報告書、あるいは、NFR を評価するためのメトリクスは、これまでも数多く提案されている[1] [2] [3] [4] [6] [7] [11]。本論文で提案する方法は、それら既存のガイドライン、報告書、メトリクスを基盤として、RFP に記述すべき NFR を、より委託元企業 (ユーザ) の視点で評価する手順を示すものである。具体的には、評価対象を、ユーザにとって重要度の高い「保守と運用に関する 55 個の非機能要件」に限定した上で、要件記述の明確さを最大 5 段階で評価するためのメトリクス (評価基準スキーム) を定義し、評価結果は、RFP の「総合評価点」と要件毎の評価点を俯瞰するための「レーダチャート」として示すものとする。

以降、2.2 節では、関連研究として、NFR に関する代表的なガイドライン、報告書、メトリクスを紹介する。3 節では、提案法を示し、4 節では、WWW 上に公開されていた 29 件の RFP を対象としたケーススタディの結果を示し、提案法

の適用容易性や有用性について議論する。最後に、5 節では、まとめと今後の課題について述べる。

2.2 関連研究

2.2.1 非機能要件に関するガイドライン

日本ユーザ情報システム協会 (JUAS) による「非機能要求仕様定義ガイドライン」[3]には、ソフトウェアライフサイクルを通じて使用することが推奨される 200 個を超える非機能要件が、ISO/IEC09126 等に準拠する形で示されている。ただし、ソフトウェア開発終了後の保守や運用に関する非機能要件は、必ずしも網羅されていない。一方、「システム構築のトラブルを回避するための IT システム契約締結の手順とポイント」[7]、および、「情報システム調達のための技術参照モデル (TRM)」[11]は、ユーザとベンダ間でソフトウェア開発契約を締結する上で重要となる、サービスレベルに関する合意 (Service Level Agreement: SLA) に必要な要件を示すとともに、保守と運用に関する非機能要件も数多く示されている。提案法では、これら 3 つのガイドラインで示された非機能要件を、評価対象の候補とする。

「システム/ソフトウェア製品の品質要求定義と品質評価のためのメトリクスに関する調査報告書」[6]には、利用者ニーズに応えるソフトウェア品質の確立、および、そのために広く利用可能なメトリクスの選定を目的とする事例調査の結果がまとめられている。報告には、非機能要件の重要度に関するユーザ・ベンダ企業へのアンケート結果が含まれている。提案法では、このアンケート結果を、評価対象とする非機能要件の選定に利用する。

多種多様な非機能要件間の関係を明らかにする研究も行われている。日本情報システム・ユーザー協会 (JUAS) による「ソフトウェア開発管理基準に関する調査報告書」[4]では、品質目標 (SLA 指標)、運用容易性、障害対策、災害対策といった観点で、非機能要件が整理されている。また、情報処理推進機構ソフトウェア・エンジニアリング・センター (IPA-SEC) による「共通フレーム 2007」[1]では、運用と保守のプロセスに関する非機能要件の整理がなされている。提

案法では、これら2つの成果に基づき、評価対象とする非機能要件 55 個を 3 階層でグループ化している。

2.2.2 RFP や非機能要件に非機能要件に関する従来研究

IPA-SEC による「非機能要求グレード」[2]は、情報システムにおけるセキュリティや性能、業務の手順など、機能以外に関する要件(非機能要件)を定義すると共に、要件に対する要求レベルを評価し、ユーザ・ベンダ間で合意を形成するための枠組みを与えるものである。要件を階層的にグループ化し、評価基準を要件毎に定義するというアプローチは、提案法と同じであるが、要求レベルの評価はベンダ視点で行われ、ユーザにとって重要な保守に関する要件などについては言及されていない。

RFP や要求仕様書など、ソフトウェア開発の初期に作成される文書の評価に、自然言語処理技術を用いる研究も報告されている。佐藤らは、要求仕様における品質要求の含有率を、形態素解析に基づく重要語句の抽出などにより測定する具体的な方法とツールを提案している[10]。評価対象には非機能要件も含まれているが、評価の粒度は、「セキュリティ」、「成熟性」、「運用性」などであり、提案法に比べると大きい。

2.3 提案方法

2.3.1 概要

提案法は、ソフトウェア開発に向けて作成される提案依頼書(Request For Proposal: RFP)の品質を定量的に評価するものである。品質評価の観点は、「運用と保守に関する非機能要件」に関する記述の有無、および、明確さである。評価結果は、RFP の総合評価点(100 点満点)、および、要件毎の評価点を俯瞰するためのレーダチャートとして示される(図 2.1 参照)。

提案法の主な利用者は、RFP 作成者(ソフトウェア開発をベンダに依頼するユーザ)である。RFP 作成者は、ベンダに対する RFP の提示に先立ち、非機能要件に関する記述の明確さを提案法により定量的・視覚的に把握する。明確

に記述されていない要件があれば、必要な加筆修正を RFP に対して行う。

2.3.2 評価対象とする非機能要件

評価対象とするのは、2.1 で示した 3 つのガイドライン[3][7][11]で示されている非機能要件のうち、保守と運用に関する 55 個の非機能要件である。これは、本提案法の主な利用者となる委託元企業(ユーザ)が、ソフトウェアと最も直接的に関わるのが「保守と運用」であり、それら要件をベンダに正確に伝えることが RFP 作成の主要な目的のひとつと考えられるからである。また、非機能要件は、セキュリティ対策、冗長化、応答時間といったアーキテクチャの制約条件となる場合が多く、アーキテクチャの実現可能性を評価する上でも役立つ。これとは反対に、ベンダによるソフトウェア開発管理に関する要件、ユーザが自身のために行う開発管理に関する要件(ベンダに伝える必要性の低い要件)は、評価対象とはしていない。

ユーザ企業への調査結果[6]から、実際にユーザが使用しているあるいは使用したいとしたメトリクスの中から、RFP に記述すべきメトリクスを回答数の多いものからランキング付けし、38 個を抽出した。しかし、RFP に記述すべき「保守・運用要件」の網羅性を高めるためにユーザ企業の調査でアンケート対象となっていない非機能要件については、文献[3][7][11]に基づいて、サービスレベルに関する合意に必要な非機能要件を選定し(17 個)55 個の要件を評価対象とした。55 個の要件のうち 34 個は運用に関する要件、21 個は保守に関する要件である。また、55 個の要件のうち 17 個は、サービスレベルの合意に必要な要件である。残る 38 個は、文献[6]で実施されたアンケートにおいて、3 分の 1 以上のユーザ企業が、「RFP に実際に記述している」あるいは「記述すべき」と回答した要件である。

2.3.3 非機能要件評価シート

非機能要件評価シートは、評価対象とする 55 個の非機能要件それぞれについて、「評価メトリクス(明確さの評価基準スキーム)」と「重要度(評価における重み)」を与えるものである(図 2.1 参照)。なお、評価対象とする要件が 55 個

と多数にのぼるため、評価結果の俯瞰が難しくなる可能性がある。そこで、類似する要件をグループ化し、17 個の「中項目」として設定し、更に、それら中項目を、ソフトウェア利用者の観点で設定した 7 個の「大項目」に対応付けている。

評価対象とする要件それぞれについての記述内容は次の通りである。

■非機能要件 i

名称:

定義:

■評価メトリクス(評価点 s_i)

明確さ4の評価基準

3の評価基準

2の評価基準

1の評価基準

0の評価基準

■重要度 w_i

提案法では、各要件は最大 5 段階で評価される。評価点の取りうる値は、0 から 4 の整数値である。「明確さ評価基準」は、文字通り、当該要件の明確さを評価するための基準を示すものである。当該要件が(十分に)明確に記述されている場合の評価点は 4、記述がない、もしくは、記述の明確さが著しく低い場合は 0 となる。ただし、要件によっては、記述の明確さに区別はなく記述の有無だけで評価できる要件、記述の明確さについての議論や検討が(現時点では)十分ではなく 5 段階評価が難しい要件、などがある。そうした要件については、明確さ 3 の評価基準、同 2 の評価基準、同 1 の評価基準のいずれか、もしくは、全てを「該当なし(N/A)」とできるものとする。例として、いくつかの非機能要件とその明確さ評価基準を図 2.2 示す。図 2.3(a)に示す非機能要件「バックアップ方式」では、5つ全ての評価基準が示されており、5 段階評価が行われる。図 2.3(b)に示す非機能要件「システムソフト」では、明確さ 3 と 1 の評価基準評価が「該当なし(N/A)」となっており、3 段階評価となる。図 2.3(c)に示す非機能

要件「応答時間」では、明確さ 3 から 1 の評価基準が全て「該当なし(N/A)」となっており、2 段階評価となる。

「重要度」は、RFP における当該要件の重要度を相対的に示す数値である。前述の通り、要件の明確さの評価点を取り得る値は、全ての要件において、0 から 4 の整数値である。そこで、RFP の総合評価点(100 点満点)の算出において、複数の要件の評価点を加算するにあたって、この重要度を重みとして用いる。要件の重要度は、対象ソフトウェアのドメインや利用組織毎に異なり、一律に定めることは出来ない。本論文では、一例として、文献[6]で実施されたアンケートにおいて、「重要な要件であり、RFP に実際に記述している」あるいは「記述すべき」と回答したユーザ企業数に基づき重要度を決定した。例えば、「バッチ処理正常終了率」の重要度は「オンラインシステム稼働率」の重要度の 6.2 倍となっているが、これは、同アンケートにおいて、上記のように回答したユーザ企業数が 6.2 倍あったことを意味する。同アンケートの対象外の要件については、システム発注・開発に長年携わってきたエキスパートの意見に基づき重要度を決定した。その上で、評価対象とする 55 個の非機能要件全体で、重要度(重み)の合計が 100 となるよう正規化を行った。その結果、重要度が最も高い要件は「バッチ処理正常終了率」で重要度は 6.2、最も低い要件は「オンラインシステム稼働率」、「アクセス監査」など 18 個の要件で重要度は 1.0 となった。

非機能要件評価シート(抜粋)



図 2.1 提案する RFP 評価法の概要

非機能要件 No.27	
名称	バックアップ方式
定義	データ及びハードウェアに関するバックアップ仕様
評価マトリクス	
明確さ4の評価基準	ハードウェア及びソフトウェアのバックアップ構成が系統的に記述されている。
明確さ3の評価基準	ハード及びソフトのバックアップについて記述されている。
明確さ2の評価基準	バックアップの記述はあるが具体的な方式の記述がない。
明確さ1の評価基準	バックアップ方式について提案を要求している。
明確さ0の評価基準	バックアップ方式についての記述がない。
重要度	3.6
カテゴリ	大項目: 障害対策 中項目: 冗長化

(a) 5段階評価

非機能要件 No.41	
名称	システムソフト
定義	システムで使用するOS及びユーティリティソフトウェア
評価マトリクス	
明確さ4の評価基準	使用するシステムソフトウェアの名称が具体的に記述されている。
明確さ3の評価基準	N/A
明確さ2の評価基準	使用するシステムソフトウェアの名称が具体的に記述されていない。
明確さ1の評価基準	N/A
明確さ0の評価基準	システムソフトウェアについての記述がない。
重要度	1.5
カテゴリ	大項目: 保守生産性 中項目: 保守容易性

(b) 3段階評価

非機能要件 No.9	
名称	応答時間
定義	システムとしての応答時間(画面操作時のデータ更新、通信時間など)
評価マトリクス	
明確さ4の評価基準	応答時間が目標時間として(数値で)記述されている。
明確さ3の評価基準	N/A
明確さ2の評価基準	N/A
明確さ1の評価基準	N/A
明確さ0の評価基準	応答時間の目標時間が記述されていない。
重要度	1.3
カテゴリ	大項目: システム運用評価 中項目: 稼働品質性能

(c) 2段階評価

図 2.2 明確さ評価基準の例

2.3.4 評価結果

「非機能要件評価シート」に基づく評価結果は、RFP の「総合評価点」と要件毎の評価点を俯瞰するための「レーダチャート」に大別される。総合評価点 S は、評価対象とする 55 個の非機能要件それぞれに対する評価点を、その重要度で重み付けした加重和である。

$$S = \sum w_i s_i / 4 \quad (i = 1, \dots, 55)$$

ここで、 s_i は、要件 i の評価点、 w_i は要件 i の重要度である。55 個の非機能要件全てが明確に記述されている場合、総合評価点 S の値は 100 となり、記述に明確さが無い、あるいは、記述そのものがないほど、要件の重要度に応じて減点されていることになる。

レーダチャートは、要件間での評価点の比較などが容易に行える表現形式である。ただし、提案法では、評価対象とする非機能要件が 55 個と多数にのぼるため、それら全ての評価値をレーダチャートで表現することは現実的ではない。そこで、「非機能要件評価シート」において設定した「大項目」および「中項目」を単位としてレーダチャートを作成する(図 2.1 参照)。「大項目レーダチャート」では、大項目それぞれに属する要件の評価点の平均値を示す。「中項目レーダチャート」でも、同じく、中項目それぞれに属する要件の評価点の平均値を示す。平均値が取り得る値は、いずれも、0~4 であり、要件が明確に記述されているほど高い値となる。

2.4 ケーススタディ

2.4.1 概要

提案法の適用容易性や有用性を評価するために行ったケーススタディの結果について述べる。ケーススタディでは、地方自治体、図書館、政府機関、大学、病院などが、ベンダ候補企業向けの入札情報として WWW 上に公開していた 29 件の RFP を評価対象とした。RFP の評価は、各 RFP の作成者ではなく、システム発注・開発に 10 年以上携わってきたエキスパート1名が、対象

RFP 全てに対して行った。

RFP の評価に要した時間は、RFP1件あたり最大 1 時間程度であった。評価者は、対象 RFP で表されるシステムやそのドメインに関する知識を十分に有していたわけではなかった。しかし、対象 RFP を熟読することで、非機能要件 55 項目それぞれの評価点を支障なく決定することが出来た。RFP 作成者自身であれば、より短い時間で評価が可能であることは容易に推察される。

また、提案方法は RFP のみに基づいて実施可能であり、対象 RFP を公開している団体や RFP 作成者に対してインタビューを行ったり、追加資料を求めたりする必要のないことも確認された。このことは、(RFP 作成者自身を含む)複数人で RFP を評価し、デルファイ法などにより、より客観性・妥当性の高い結果を得ることが、比較的容易であることを意味する。

2.4.2 総合評価点

図 2.3(a)は、29 件の RFP の総合評価点の分布を、RFP が表す情報システムの 5 つのドメイン毎に示した箱ひげ図である。5 つのドメインとそれぞれの RFP 件数は次のとおりである。

地方自治体 6 件

図書 8 件

政府機関 5 件

大学 5 件

病院 5 件

箱ひげ図は、データ分布の様相を視覚的にとらえやすく表すために工夫された図である。箱の中に引かれた横線がその分布の中央値を、箱の下辺と上辺がそれぞれ第一四分位数、第三四分位数を、更に、上下にのびたヒゲの先端が、それぞれ最大値と最小値を表す。なお、外れ値がある場合は、箱やひげとは別に、○印で表される。

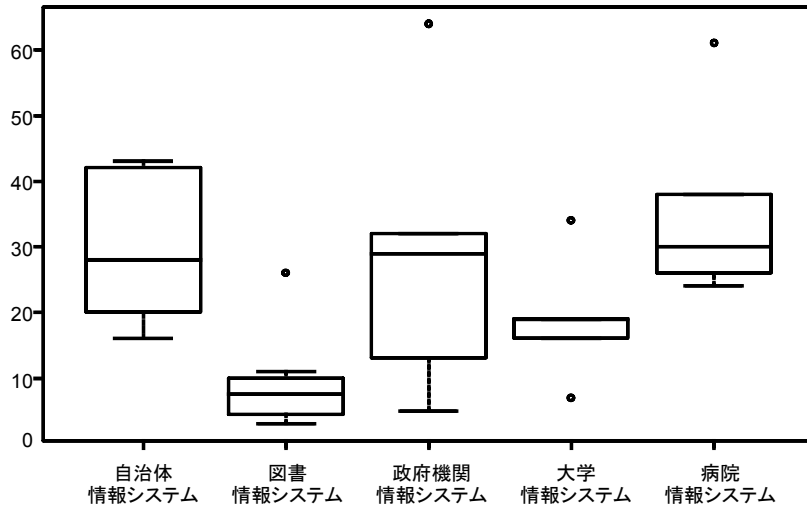
図 2.3(a)より、ドメインによって総合評価点に大きな違いのあることが分かる。また、総合評価点が 60 点以上となったのは、政府情報システムと病院情報システムのそれぞれで 1 件のみである。提案法では、評価対象とする 55 個の非

機能要件全てが RFP において明確に記述されているべき、という立場で評価が行われている。総合評価点は、満点となる 100 点にできるだけ近いことが望まれる。しかし、大半の RFP は総合評価点が 100 点からほど遠く、非機能要件がまだまだ明確には記述されていない、ということになる。特に、図書館情報システムでは、総合評価点の中央値が 10 点未満であり、RFP に改善の余地が大きく残されていると言える。

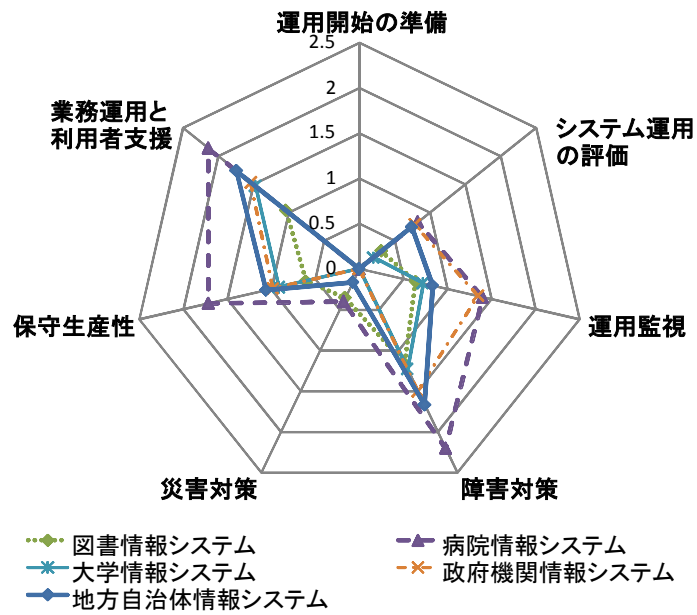
2.4.3 レーダチャート

大項目と中項目の評価結果となるレーダチャートを図 2.3(b)(c)にそれぞれ示す。同図では、5 つのドメインそれぞれにおける評価点の平均が示されている。図 2.3(b)を見ると、5 つのドメイン全てにおいて大項目「運用開始の準備」の評価点が 0、「災害対策」が 0.5 以下、「システム運用の評価」が 1.0 以下と極めて低いことが分かる。評価点が 0 となった「運用開始の準備」は、図 1 に示すとおり、3 つの非機能要件「運用移行許容障害発生率」、「テスト密度」、「テストカバレッジ」で構成されている。評価点が 0 ということは、これらが全て RFP に一切記述されていなかったことになる。必要がないから記述されていなかったとも考えられるが、「非機能要件を十分に提示している」とするユーザ企業が 22.6%に過ぎないとの調査結果[4]もあることから、ここでは、「必要だが記述されていなかった」との立場をとる。今回のケーススタディにおけるユーザは、地方自治体、政府機関、大学、病院等であり、情報システム部門を持たず、テストに関する知識や経験が不足していた可能性がある。その結果、テストに関連する要件が記述されず、評価点が 0 となったと推察する。

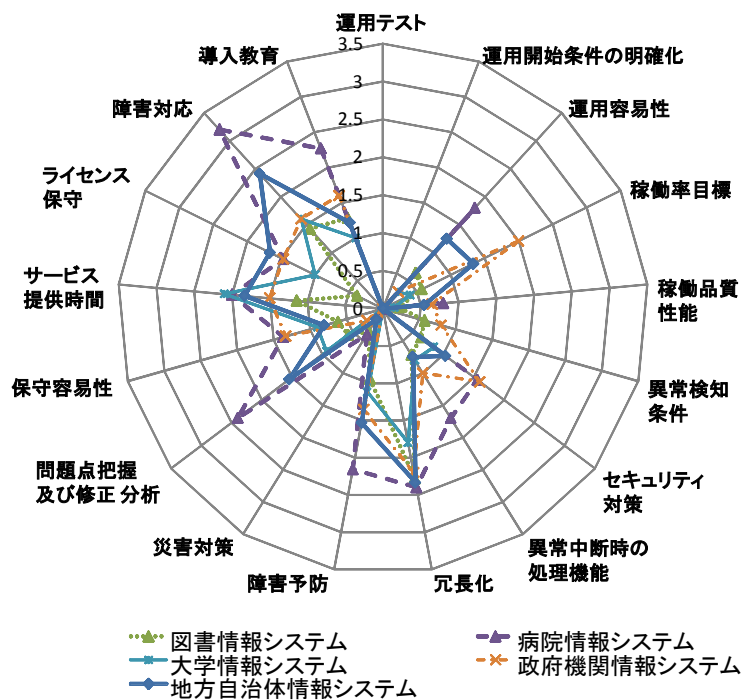
評価点が 1.0 以下となった「システム運用評価」は、同じく図 1.1 に示すとおり、3 つの中項目「運用容易性」、「稼働率目標」、「稼働品質性能」で構成されている。図 2.3(c)によれば、このうち、「稼働品質性能」の評価点がどの分野においても低いことが分かる。「稼働品質性能」は 11 個の非機能要件で構成されており、更に詳細な評価・分析が可能であるが、ここでは省略する。詳しくは、文献[9]を参照されたい。



(a) 総合評価点



(b) レーダチャート: 大項目



(c) レーダチャート：中項目

図 2.3 システムドメイン毎の評価点

2.4.4 ベンチマーキング

ケーススタディ結果のひとつとして、提案法におけるベンチマーキングについて述べる。先にも示した通り、提案法では、評価対象とする55個の非機能要件全てが RFP において明確に記述されているべき、という立場で、いわゆる減点法により評価が行われる。RFP 作成者の目標は、総合評価点が100点、レーダチャートで示される全ての項目の評価点が4点、となる RFP を作成することと言える。

ただし、100点満点の RFP を作成することが、(現時点において)現実的であるかどうかについては議論の余地がある。提案法では、既存のガイドライン、および、RFP 作成者となるユーザ企業へのアンケート結果に基づいて、評価対象となる非機能要件を選定し、記述の明確さの評価基準や重要度等を要件毎に定めている。しかし、それら要件を明確に記述することの容易性については考慮されていない。限られた工数・期間の下では、明確に記述されにくい要件が存在する可能性もある。目標としての100点満点とは別に、標準値あるいは基

準値を設定し、個々の RFP 評価点との比較を行うベンチマーキングも必要であると考えられる。

ここでは、一例として、評価対象とした 29 個の RFP のうち、総合評価点が高かった 3 個の RFP (RFP トップ 3) における平均評価点を、各要件に対する評価点の基準値とした。なお、基準値の設定においては、特異点、あるいは、例外的と思われる値(評価点)は除外する必要がある。特に、著しく高い評価点は、目指すべき高い目標として基準値に組み入れるべきとされる一方で、特異点、あるいは、例外的として基準値設定から除外すべき場合もある。基準値設定に用いた 3 個の RFP のうち 2 個の総合評価点はおよそ 60 点で、他の RFP に比べれば著しく高い値となっている。ただし、100 点満点中の 60 点であり、要件によっては、他の RFP よりも平均評価点が低くなる場合もあることから、現時点では、特異点、あるいは、例外的とは見なさず基準値設定に用いた。図 2.4 は、総合評価点が中央値であった RFP (RFP M と呼ぶこととする) における評価点を基準値と比較した結果である。一般論で言えば、RFP M の評価値と基準値の差が大きい要件ほど、記述の明確さに改善の余地があることになる。同図より、要件「稼働率目標」、「異常検知条件」、「サービス提供時間」などが該当する。

個別の要件について、もう少し詳しく見ていくと、例えば、要件「導入教育」の評価点は、RFP M では 4 点、基準値、すなわち、RFP トップ 3 の平均では 2.89 点となっている。評価点が満点の 4 点であることから、RFP M において同要件が相対的にも絶対的にも極めて明確に記述されていることが分かる。

また、要件「運用容易性」に注目してみると、RFP M の評価点は 2 点、基準値も 2.17 点とほぼ同じである。RFP M の評価点だけで判断すると、同要件は必ずしも明確に記述されていない、ということになる。しかし、RFP トップ 3 と同程度には明確に記述されており、現時点では、改善の余地はそれほどないかもしれない。一方、RFP M において、評価点が同じ 2 点となっている要件「障害予防」について見てみると、基準値は 3.20 点となっており、より明確に記述する余地が残されていることが分かる。こうした違いは、RFP M の評価点だけを比べても分からない。他にも、要件「冗長化」について言えば、RFP M の評価点は 3 点と要件「運用容易性」よりも高い評価となっているが、基準値は 4 点であり、

要件「運用容易性」よりも既に明確に記述されてはいるが、更に明確に記述する余地が残されていることが分かる。

なお、RFP Mにおいて評価点が0点となっているのは、要件「運用テスト」、「運用開始条件の明確化」、「稼働率目標」、「稼働品質性能」、「異常検知条件」、「災害対策」、「ライセンス保守」の7要件である。このうち、要件「運用テスト」、「運用開始条件の明確化」については、基準値も0点となっているが、いずれもユーザ企業に対するアンケート[6]において重要であるとの回答数が多い要件である。特に、高い信頼性が要求されるドメインでの委託ソフトウェア開発においては、ベンダがシステム開発の完了を確認し、ユーザが運用を開始する条件としてRFPに記述されるべき要件である。一方、残りの5つの要件については、より明確に記述する余地があり、RFP Mにおける記述の不明確さには、個別の原因や理由があると考えられるべきである。

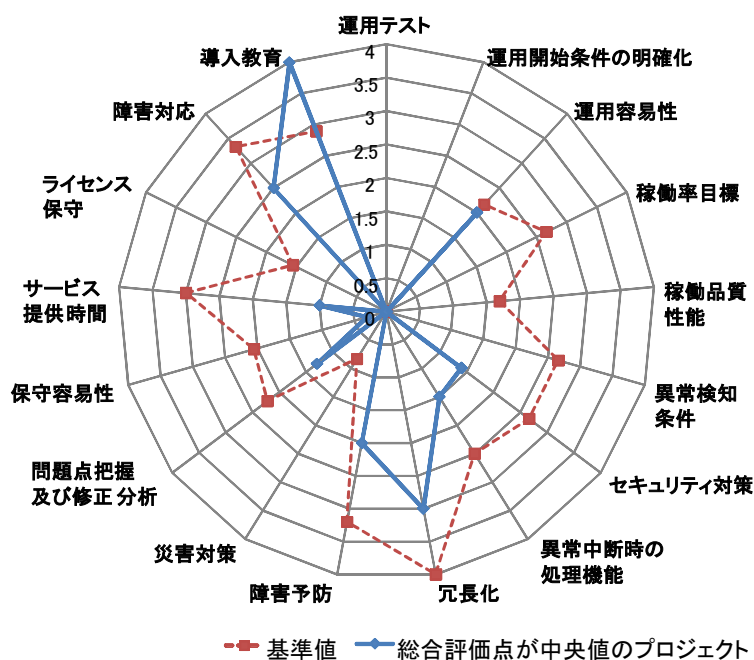


図 2.4 ケーススタディ結果：基準値（RFP トップ3）との比較

2.4.5 評価者間の評価点のばらつき

評価者間のばらつきを確認するため、評価者を2名追加し、エキスパートとの

間で評価結果を比較する実験を行った。追加した評価者のうち1名は、ソフトウェア工学を専門とする、業務経験のない大学教員(以降、教員と記す)である。もう1名は、エンタープライズ系のソフトウェアエンジニアとして20年以上の経験を有する者(以降、エンブラ系SEと記す)である。29件のRFPのうち、各ドメイン(地方自治体、図書館、政府機関、大学、病院)から各1件をランダムに選択し評価対象とした。

実験の結果、まず、各要件に対する評価点の評価者間での差(絶対値の平均)は、1非機能要件あたり、エキスパートと教員の間で0.367、エキスパートとエンブラ系SEとの間で0.585となり、1未満(5段階評価における1段階未満)となった。評価点に有意差(フリードマン検定、有意水準5%)が認められたのは、病院情報システムのRFPに対するエキスパートの評価点とエンブラ系SEの評価点のみであった。そのケースにおいて、評価点の差が特に大きかった要件は、「スループット」「最大負荷スループット」「最大停止時間」「ターンアラウンド時間」「保証期間」の5つであった。これらはいずれも、要件に関する数値情報が記述されていれば4点、されていなければ0点となる要件で、エキスパートによる評価はいずれも0点、逆に、エンブラ系SEによる評価はいずれも4点であった。実際には、これら5つの要件に関する数値情報はRFPには記述されておらず、エンブラ系SEによる評価は妥当でないことがわかった。エンブラ系SEに追加インタビューしたところ、「数値情報は示されていなかったが、要件に関する記述は見られたので4点と評価した。数値情報の有無を厳密に評価に反映しなかったのは少し寛大なのでは、と指摘されてもいたしかたない。」との回答が得られた。このことから、数値情報の有無が評価に直結する要件については、そのことを評価者に徹底することが必要であり、また、徹底することで、評価者間で評価のばらつきを小さく抑えることが期待される。

次に、総合評価点(100点満点、重み付き)については、表2.1に示す結果となった。エンブラ系SEによる病院情報システムに対する評価点を除くと、教員およびエンブラ系SEによる評価点とエキスパートによる評価点との差は、最大でも6.09にとどまった。

以上より、実務経験のない大学教員であってもエキスパートと有意差のない評

表 2.1 各評価者の各 RFP に対する総合評価点

評価者 ドメイン	エクス パート	教員	エンプラ 系 S E
地方自治体	2.71	5.99	6.88
図書館	27.31	32.79	33.40
政府機関	18.91	15.16	16.91
大学	5.06	5.88	5.34
病院	42.75	44.08	64.90

価を行えること、また、数値情報の記述が求められる非機能要件については、具体的な数値が記述されていなければ評価点は“0”とすべきことを徹底することで、評価のばらつきを抑えられる可能性があることが分った。本結果の信頼性を増すため、より多くの評価者を被験者として評価実験を行うことが今後の課題となる。

2.5 まとめ

2.5.1 評価結果と考察

本章では、ベンダへの提案依頼書(RFP)提示に先立ち、RFP 作成者であるユーザ自身が、RFP の品質を定量的に評価する方法を提案した。評価対象は、ユーザにとって重要度の高い「保守と運用に関する 55 個の非機能要件(NFR)」であり、評価の観点、その記述の明確さ、である。記述の明確さは、最大 5 段階で評価され、その結果は、RFP の「総合評価点」と要件毎の評価点を俯瞰するための「レーダチャート」として示される。地方自治体、図書館、政府機関、大学、病院などが WWW 上に公開していた 6ドメイン 29 件の RFP を評価対象としたケーススタディによって、記述が不十分な要件を特定したり、基準値との比較を通じて特に改善が必要な特性を明らかにしたりできることなどが

確認された。加えて、ドメインや要件によって評価点やそのばらつきに比較的大きな差があることが、総合評価点の比較やレーダチャートによる俯瞰により明確となり、提案法に基づく RFP ベンチマーキングの可能性についても議論を行った。なお、評価は RFP のみに基づいて実施可能であり、評価に必要な時間も、RFP1件あたり最大 1 時間程度であった。

提案法は、RFP を対象としたものであり、ベンダへの提示に先だってユーザのみが利用するものと位置づけられている。ただし、RFP に基づいて作成される技術仕様・技術提案書や契約仕様書へと適用範囲を拡げることは比較的容易である。その場合、技術仕様・技術提案書の作成においてベンダが提案法を利用する、また、契約仕様書の作成に向けた技術協議において、ユーザとベンダの双方が提案法を利用し、非機能要件に関する合意形成を効率よく行う、といったことも考えられる。

また、関連研究においても少し紹介したが、ソフトウェア開発で作成される文書の評価に、自然言語処理技術を用いる研究が盛んに行われている。提案手法においても、例えば、非機能要件記述に含まれる典型的な語句や表現を自然言語処理技術で抽出し、非機能要件の文例集を作成することが考えられる。文例集があれば、RFP や対象ドメインに関する知識が十分でない者でも、提案手法による評価が可能に、あるいは、より容易になる。評価者による評価結果のばらつきが減れば、評価法に基づく RFP ベンチマーキングの信頼性や有用性も高まる。テキストマイニングや機械学習といった技術と組み合わせることで、RFP 評価の自動化にも道を開くことになる。

2.5.2 課題

提案手法を用いて超上流工程で RFP に記述されている非機能要件を定量評価することにより、ユーザの満足度は向上することが推定される。なぜなら、不十分な記述のユーザ要求が数値として明確になり、これを明確化することにより要件定義が行われることが期待されるからである。しかし、ユーザの満足度を評価するためには、定義された非機能要件についてソフトウェアライフサイクルを通じた「要求管理」を継続することにより、要求の変更を定量評価することが求

められる。文献によれば「マイクロソフトのプロジェクトでは 30%の要求をあらかじめ抽出することはできず、開発工程の途中で追加された」[49]とされている。委託ソフトウェア開発では、ソフトウェアライフサイクルを通じて、提案手法によるユーザ要求の追加変更を定量的に評価し、顧客満足度を検証することが今後の課題である。

また、RFP に記述された非機能要件の記述内容について、多くの RFP を人手で評価を行うには時間を必要とする問題点については、多くの RFP を自動評価することにより時間を短縮し、相対的に比較評価できることが求められる。自然言語で記述された RFP の非機能要件の記述内容を自動評価するためには、非機能要件に関連する語句が RFP にどの程度記述されているかを評価することにより、明確さを評価する手法が有効であると考えられる。そこで、次章以後では、非機能要件に関連する NFR キーワードを、テキストマイニングを用いて RFP テキストから抽出し、自動評価することにより、非機能要件の記述内容の明確さを評価する手法について述べる。

第 3 章 機械学習による非機能要件の自動評価

3.1 概要

第 1 章で述べた手法[9][15]では、RFP に記載されるべき 55 個の非機能要件(図 2.1 参照)の記述をユーザ視点から評価する枠組みを提案し、ケーススタディにおいてエキスパート技術者による評価を行った。しかし、「非機能要件評価シート」に基づいた“人手による評価”(以後は手動評価と称する)は、エキスパートがいることが前提となること、および評価に時間を要することが問題となる。そこで、本章では、これらの問題を解決するために、RFP に含まれるキーワード群から非機能要件に関連するキーワード(以後 NFR キーワードとする)を抽出し、その出現頻度に基づいた機械学習による自動評価手法を提案する。

提案手法では、最初に自然言語で記述された複数の RFP から NFR キーワードをテキストマイニングにより抽出する。NFR キーワードと「中項目」の関連付けにおいて、より明確に非機能要件の記述内容を評価するために「非機能要件評価シート」の階層をさらに細分化し、粗いレベルの「中項目」と抽出した NFR キーワードの間に「小項目」(非機能要件)を設けることにより、「NFR キーワード評価シート」を作成した。次に、抽出された NFR キーワードと「小項目」にマッピングされた NFR キーワード群の出現頻度をデータとして、教師あり機械学習によりモデル化する。モデル化手法としては、分類精度が最も高く、高次元データ解析に向いており、汎化能力に優れているランダムフォレストを用いる[17]。評価実験では、NFR キーワードによる自動評価について、2 段階、3 段階及び 5 段階の手動評価の結果を教師データとした機械学習を行い、その結果について比較する。以下、3.2 節で関連研究、3.3 節で RFP と NFR キーワードの相関関係についての予備調査、3.4 節で機械学習による提案手法、3.5 節で機械学習による非機能要件評価実験、3.6 でまとめについて述べる。

3.2 関連研究

ソフトウェア開発の上流工程にて、自然言語で記述された非機能要件を評

価する国内での先行研究は少なく、特に非機能要件の記述内容の定量評価については示されていない。ソフトウェア開発の上流工程にテキストマイニングを応用する実用的な研究として、日本国内では福田[19]による概念データモデリング、SOA のサービス抽出及びサービスのクラスタ化を支援する試みが提案されているが、クラスタ分析及び文書・語句行列による評価の試行にとどまっており、定量評価は行われていない。

また、今村ら[12]による技術文書からの用語知識を自動的に獲得する実験は、共起関係としての「係り受け先」と「文節内の後方」が体系的意味分類として有効であることを指摘しており、自然言語で記述された技術文書から語句を抽出する場合に有益な手法を示唆している。しかし、機械学習によるモデル化は今後の課題としている。

さらに、自然言語で記述された文書の質判定については、英語で記述された論文を対象としたさまざまな手法による質判定の研究[13] [14]が行われている。小林らは、論文の質が如実に反映される言語項目を素性として論文の質をランダムフォレストにより2クラス分類し、分類精度を評価している。本研究とは、非機能要件を対象としている点、および、多クラス分類を行う点が異なる。

一方、海外の先行研究では、自然言語で記述された要求仕様を、ソフトウェア開発の上流工程において評価する重要性の視点から、テキストマイニング技術により抽出した語句にもとづいて要求分析を行い、ソフトウェア設計を支援する手法及び評価に関する先行研究[19] [20] [21] [22] [23] [24]がある。本研究の対象である自然言語で記述された非機能要件に着目した先行研究としては、1)サポートベクタマシーンによる非機能要件、特にアーキテクチャに関する分類器に関する先行研究[25]、2)セキュリティ、パフォーマンス、操作性などの非機能要件に関する語句をキーワードとして、独自の分類器を用いて評価する手法を提案している先行研究[26] [27]、及び3)コンポーネントベース開発を対象として非機能要件メトリクスにもとづいた品質評価に関する先行研究[28]がある。しかし、これらの先行研究は、非機能要件をパフォーマンス、セキュリティ、可用性などの非常に大きなカテゴリにおける評価にとどまっており、評価実験におけるデータ数も少ない。一方、本研究では、より細分化された非機能要件(例

えば、セキュリティに関して言えば、アクセス権限、ウイルス対策、セキュリティ管理レベル、パスワード管理、暗号処理、情報漏洩対策、認証機能、不正アクセスなどをそれぞれ評価することを目的としている。また、評価に用いた 70 件の RFP によるモデル構築用のデータを Web にて公開することで、一般企業における利用を可能としている[16].

3.3 非機能要件キーワードの出現頻度予備分析

本節では、収集した RFP テキスト群を用いて「小項目」にマッピングされた NFR キーワードで記述の明確さを評価することの妥当性について予備分析を行う。機械学習を行う前処理として、収集した RFP 群の各 RFP を文書として結合した RFP テキストを作成する。RFP 群はウェブに公開されている「調達仕様書」、「要求仕様」などからさまざまな情報化対象ドメイン(第 1 章と同様に「図書情報システム」、「大学情報システム」、「医療情報システム」、「政府機関情報システム」などの公的機関における情報化システム)を対象とした。理由として、RFP 情報が公開されており入手が容易であること、および委託ソフトウェア開発による情報化システムの非機能要件が記述されていることがあげられる。

3.3.1 NFR キーワードの抽出

自動評価を行うためにウェブに公開されている 70 件の RFP 群を対象とした。詳細は 3.4 節を参照のこと。

機械学習に使用するキーワードについては、「出現頻度で上位 10% の単語しか使わなくても分類器の性能は下がらないという実験結果がある。大多数の単語の文書頻度はきわめて低く、実際には上位 10% がすべての低中頻度を含んでいるからである」[32]とされている。ランダムフォレストでは、説明変数総数の平方根(デフォルト値)の説明変数により決定木を生成し分類が行われる。評価対象とした RFP 群の中で NFR キーワードの出現頻度が高い大項目「運用要件」の「小項目」に関連付けられた NFR キーワードについて説明変数の重要度と出現頻度を表 3.1 に示す。表 3.1 によると、出現頻度で上位 10% の

NFR キーワードの場合は、説明変数の重要度は 13.16%から 90.2%であり、個別の NFR により大きな差がある。しかし、出現頻度で上位 10%より多い出現頻度上位 10 個の NFR キーワードの場合は、60.2%から 100% (ただし、キーワード数が 10 個に満たない NFR があるので 100%となっている)となることが認められる。「小項目」の NFR キーワード数が 30 個程度であれば、上位 10 個の NFR キーワードによる説明変数の重要度が 70%以上であることが分かる。よって、表 3.1 によると、さらに多くの NFR キーワードを採用すれば NFR キーワードによる重要度比率は高くなると言える。しかし、個々の NFR による違いがあるにせよ、「小項目」に対応付けられた 30 個程度の NFR キーワードであれば NFR キーワードの出現頻度が上位 10 個の重要度比率が70%以上であり、自動評価に十分な分類精度が得られると言える。

表3.1 NFRキーワードの重要度と出現頻度

NFR	説明変数	合計	出現頻度 10%	出現頻度 TOP5	出現頻度 TOP10
システム管理	重要度	16.82767	9.825959	10.69791	13.4981
NFRキーワード数=31	重要度比率	100%	58.39%	63.57%	80.21%
運用管理	重要度	24.07378	8.193873	13.74518	18.86369
NFRキーワード数=22	重要度比率	100%	34.04%	57.10%	78.36%
アクセス権限	重要度	18.40352	4.214255	6.325919	13.60273
NFRキーワード数=25	重要度比率	100%	22.90%	34.37%	73.91%
ウイルス対策	重要度	16.0936	10.10849	10.91389	11.87168
NFRキーワード数=30	重要度比率	100%	62.81%	67.82%	73.77%
セキュリティ管理レベル	重要度	16.0434	6.133548	7.626645	9.657985
NFRキーワード数=32	重要度比率	100%	38.23%	47.54%	60.20%
セキュリティ対応	重要度	19.59131	8.064515	9.119925	13.4673
NFRキーワード数=35	重要度比率	100%	41.16%	46.55%	68.74%
パスワード管理	重要度	12.40575	1.884423	6.280194	9.17485
NFRキーワード数=24	重要度比率	100%	15%	51%	74%
暗号処理	重要度	12.98212	11.68589	12.5504	12.5504
NFRキーワード数=13	重要度比率	100%	90.02%	96.67%	96.67%
情報漏洩対策	重要度	10.43894	1.374139	9.076611	10.43894
NFRキーワード数=9	重要度比率	100%	13.16%	86.95%	100.00%
認証機能	重要度	13.36928	2.699545	5.296596	9.078904
NFRキーワード数=35	重要度比率	100%	20.19%	39.62%	67.91%
不正アクセス	重要度	22.05132	8.342224	12.83992	22.05132
NFRキーワード数=16	重要度比率	100%	37.83%	58.23%	100.00%
システム異常検知	重要度	16.87974	6.509435	8.942516	13.71657
NFRキーワード数=31	重要度比率	100%	38.56%	52.98%	81.26%

3.3.2 「小項目」と NFR キーワードの関連付け

70 件の RFP から抽出した NFR キーワードについて、「非機能要件評価シート」の「非機能要件」項目と「同意概念」となる場合があることが分かった。さらに、NFR キーワード群をカテゴライズすることにより、非機能要件としての「上位概念」が明確になり、非機能要件の記述内容の明確さを評価する上で有効であると考え、「非機能要件評価シート」の「中項目」の下層に「小項目」を追加した 4 層構成となる「NFR キーワード評価シート」を作成した。「NFR キーワード評価シート」の概要を図 3.1(詳細は付録 A. を参照)に示す。また、各「小項目」に関連付けられた NFR キーワードの出現頻度の合計を図 3.2 に示す。

図 3.2 によると評価対象とした RFP 群の NFR キーワードの出現頻度が、各「小項目」により明確な差として表されることがわかる。このことは、どの「小項目」の記述の明確さが不足しているか、あるいは十分であるかについて比較し、判断することが可能であることを示唆している。次項 3.3.3 では、「小項目」と NFR キーワード出現頻度についての予備調査を行った結果を示す。

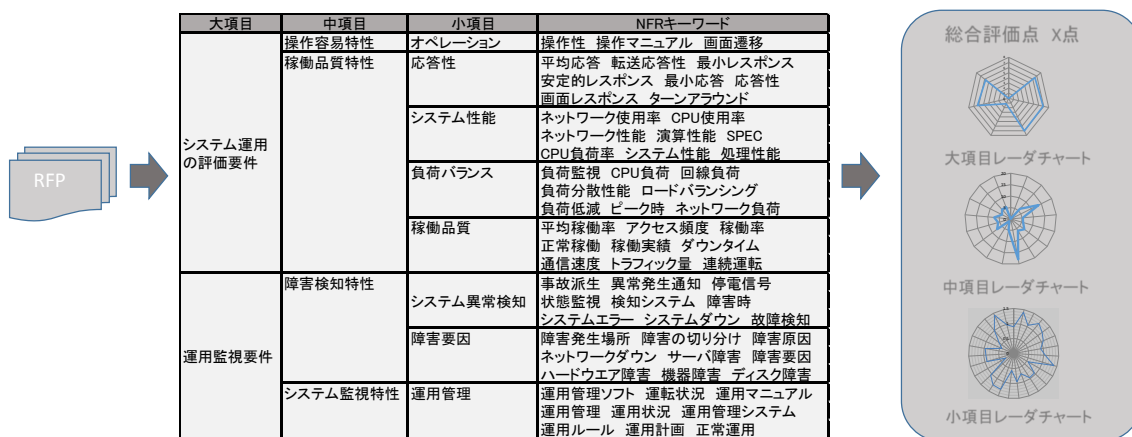


図 3.1 NFR キーワード評価シート(抜粋)

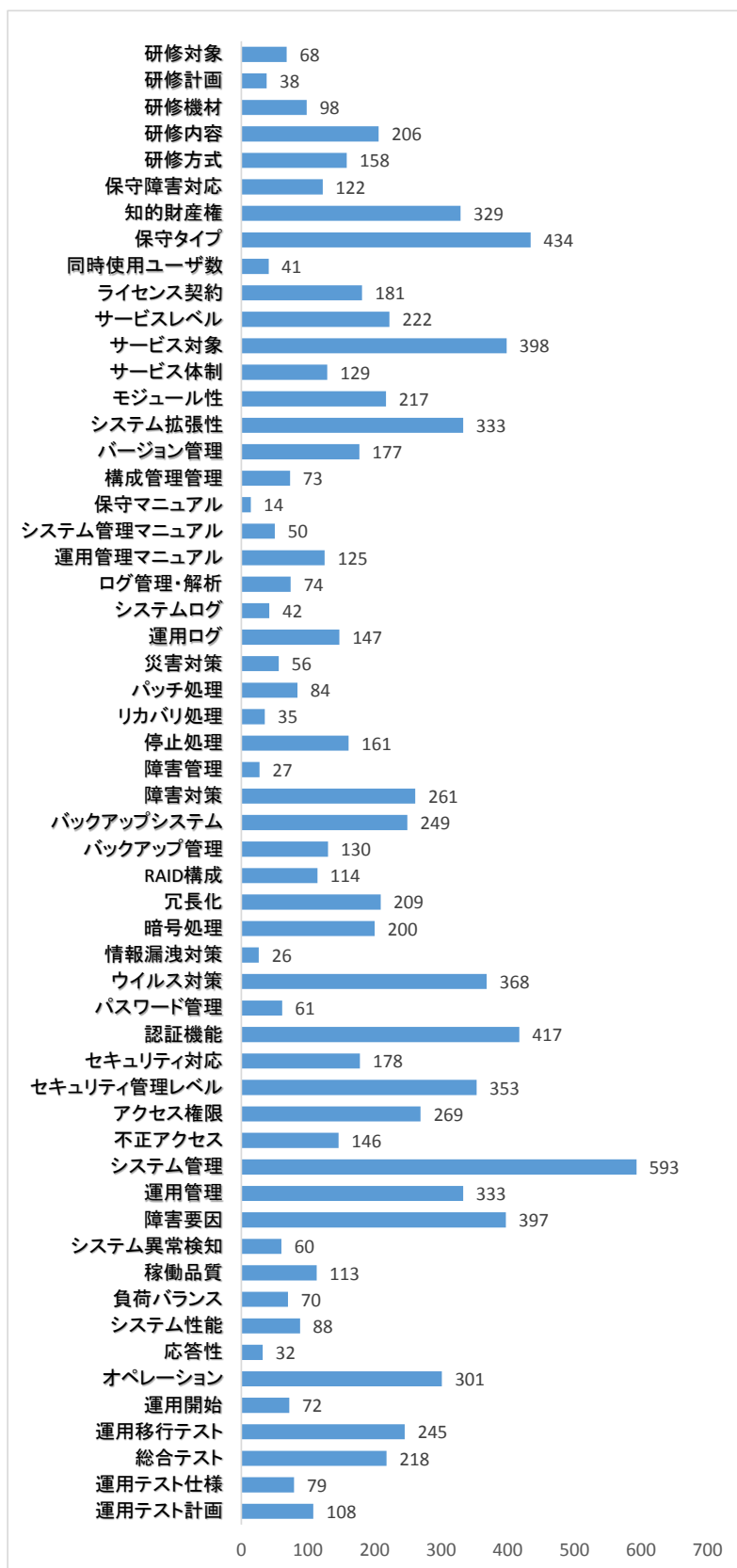


図3.2 「小項目」毎のNFRキーワード出現頻度

3.3.3 「小項目」と NFR キーワード出現頻度

NFR キーワード 776 個の内、70 件の RFP 群に出現したかどうかをカウントした RFP 出現頻度が多い 20 個の NFR キーワードを表 3.2 示す。表 3.2 よると「小項目」の内「運用管理」、「システム拡張性」、および「オペレーション」が多いことが認められる。また、収集した RFP の多くは公的機関が入札を目的にウェブ上に公開したものであり、「バージョン管理」、「保守障害対応」および「障害要因」あるいは「障害対策」に関連する NFR キーワードについて、多くの RFP に出現していることが特徴として考えられる。このことは、「小項目」に関連付けられる NFR キーワードによる評価の可能性を示唆していると言える。

表 3.2 出現した RFP 数が多い上位 20 個の NFR キーワード

NFRキーワード	RFP出現頻度	小項目
システム管理	46	システム管理
システム運用	43	運用管理
バージョンアップ	40	バージョン管理
拡張性	40	システム拡張性
障害対応	40	保守障害対応
障害発生	39	障害要因
運用管理	36	運用管理
操作マニュアル	36	オペレーション
セキュリティ対策	36	セキュリティ管理レベル
障害時	34	障害対策
保守体制	33	サービス体制
著作権	29	知的財産権
操作研修	29	研修内容
ハードウェア構成	28	システム拡張性
冗長化	25	冗長化
ネットワーク構成	25	システム拡張性
操作方法	24	オペレーション
暗号化	23	暗号処理
運用支援	23	運用管理
運用マニュアル	23	運用管理

次に、RFP に記述されている語数と NFR キーワードの出現頻度について図 3.3 に示す。図 3.3 によると、2 件の RFP については RFP 総語数および NFR

キーワードの出現回数が外れ値となっているが、残りについては RFP 総語数と NFR キーワード出現頻度に相関関係(表 3.3 参照)があり, RFP 総語数が多い場合には NFR キーワードの出現頻度も高いことが認められる. また, 外れ値の 2 件は独立行政法人の共通情報基盤システムと新設の総合医療センターに関する情報化システムであり, 評価対象とする RFP 群の中では相対的に多くの非機能要件が明確に記述されている RFP と言える.

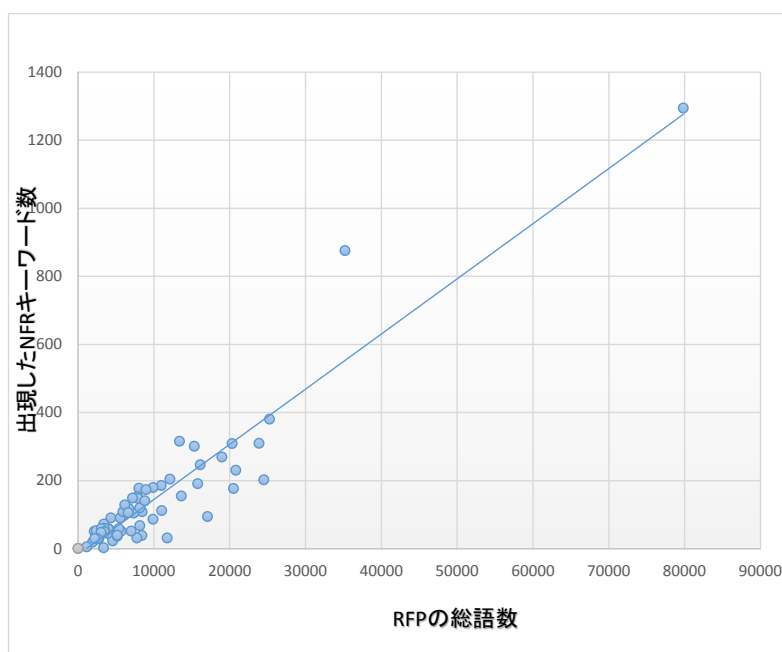


図 3.3 RFP 総語数と NFR キーワード出現頻度

さらに, 図 3.4 に RFP 毎の NFR キーワード出現頻度と「小項目」について, RFP を手動評価した総合評価ポイントを示す. 図 3.4 によると, NFR キーワード出現頻度と RFP の手動評価の相関係数は 0.649 であり, 相関があると言えるが, 強い相関があるとは言えない. これは, 一部の RFP について NFR キーワード出現頻度は低いが高評価が高いものがあることによると考えられる. 特に, 外れ値となっている RFP は医療情報システムに関するものであり, 手動評価が高い理由として, システムの応答あるいは処理時間について数値情報で記述されており, また非機能要件に関連する専門用語が複合語ではなく平易な表現で記述されている等がある. 手動評価では, 記述内容の前後関係ある

いは明確な数値で記述されている場合には、評価を高くするが、NFR キーワードとして抽出した場合には、出現頻度が低い結果となっていることが考えられる。

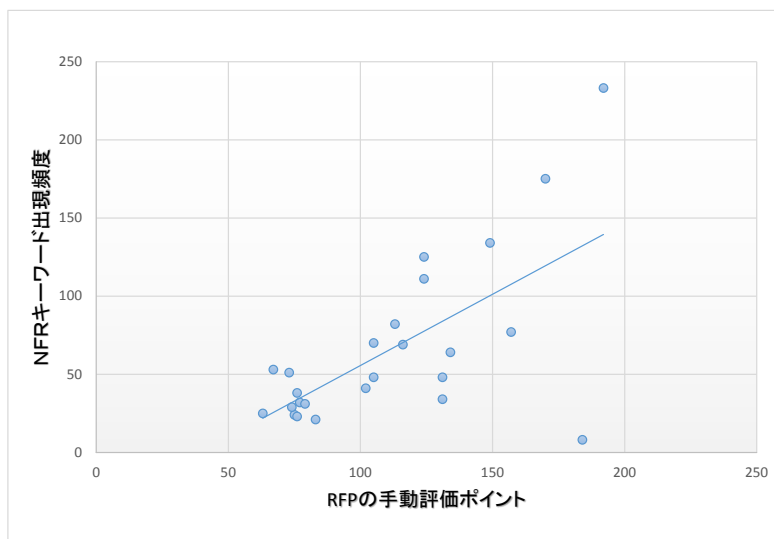


図 3.4 NFR キーワード出現頻度と RFP の手動評価

最後に、図 3.5 に NFR キーワード出現頻度と「小項目」の相関を示す。図 3.5 によると、これらの相関係数は 0.857 で、強い相関があり、NFR キーワードの出現頻度が多ければ RFP に記述されている非機能要件を「小項目」により評価することが有効であることを示している。

以上の予備調査の結果により、「小項目」にマッピングされた NFR キーワードと RFP に相関があり、NFR キーワードの出現頻度により非機能要件の記述の明確さを評価することの有効性が示されたと考えられる。そこで、次節の 3.4 節以降では NFR キーワードの出現頻度により自動評価手法を提案し、機械学習による評価実験の結果と手動評価との比較について述べる。

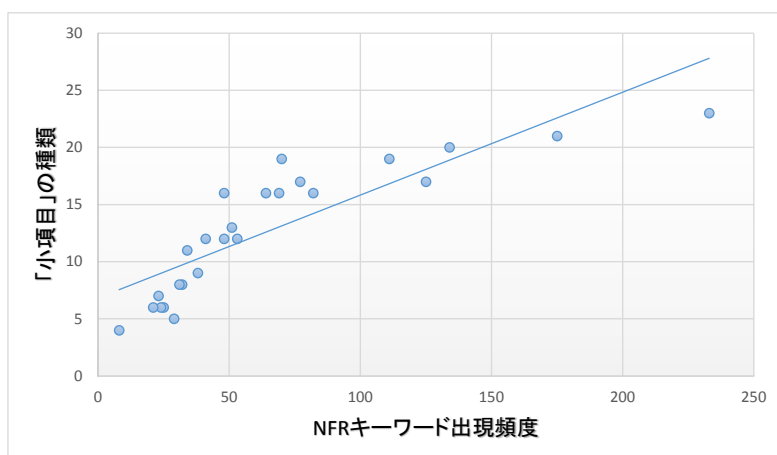


図 3.5 NFR キーワード出現頻度と「小項目」の種類

表 3.3 NFR キーワードと RFP の相関関係

相関関係データ	相関係数	相関
RFP総語数とNFRキーワード出現頻度	0.851	強い相関あり
NFRキーワード出現頻度と手動評価	0.649	相関あり
RFP毎の「小項目」の種類とNFRキーワード出現頻度	0.857	強い相関あり

3.4 提案手法

本節では、「NFR キーワード評価シート」の重要な中項目 10 個に属する 26 個の「小項目」に絞ってモデルを構築し実験を行う。モデル構築にはエキスパートの労力が必要となるが、ひとたびモデルができれば、エキスパートがいない状況においても、RFP の評価を支援できる。また、モデル構築用に、70 件の RFP のキーワード群、各キーワードの出現頻度、教師信号をまとめたデータを Web にて公開しており[16]、エキスパートがいない場合でもモデル構築・利用を可能としている。

3.4.1 「小項目」の自動評価モデル構築の概要

本項で述べる自動評価モデルは、第 2 章で作成した「非機能要件評価シート」と、テキストマイニングにより抽出した NFR キーワードをもとに作成した「NFR キーワード評価シート」の「小項目」を NFR キーワードの出現頻度により自動評価するもので

ある。ここでは、自動評価モデル構築の前処理から評価モデル構築までの手順概要について述べ、次項 3.4.2 でその詳細を述べる。図 3.6 に「小項目」の自動評価モデル構築手順の概要を示す。図中の各手順は次のとおりである。

手順1:RFP 「小項目」の手動評価

RFP に記述されている「小項目」の内容を判断し、あらかじめ評価者が「小項目」について段階評価を行う。この評価結果は、教師信号としてモデル構築用に用いられる。

手順2:RFP の形態素解析

RFP 群のテキストデータから形態素解析を行い、名詞および複合語を抽出する。

手順3:NFR キーワードの抽出

形態素解析により抽出された名詞及び複合語の中から評価対象とする「小項目」に関連する NFR キーワードを手作業で抽出し、「小項目」とのマッピングを行う。

手順4:評価モデルの構築

NFR キーワードを説明変数とし、手順1で行った手動段階評価点を教師データとした教師あり機械学習用のデータを生成する。

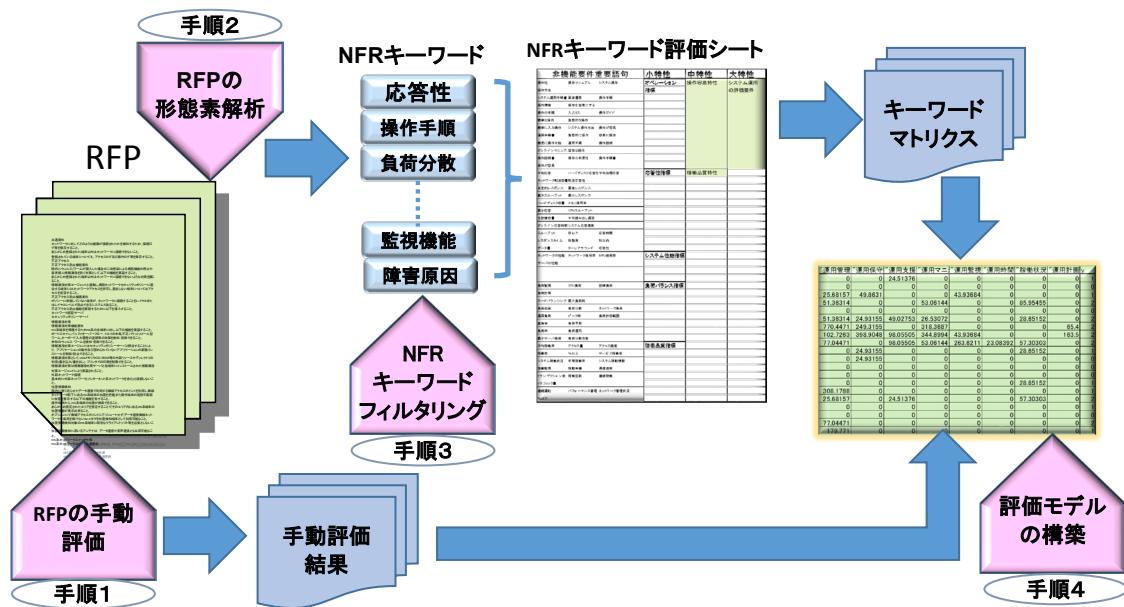


図3.6 機械学習によるNFRキーワード評価モデル構築手順

3.4.2 非機能要件評価モデルの構築

手順1：「小項目」の手动評価

エキスパートによる RFP の手动評価の結果は、RFP の機械学習による自動評価において教師データとして使用し、比較のために「小項目」ごとに2段階評価、3段階評価及び5段階評価により評価ポイントとして与えられる。これらの段階評価は、RFP の「小項目」の記述内容の明確さを評価するために記述されているか記述されていないかを評価するスケールとして、2段階評価、非機能要件の記述はあるが内容としてあいまい性を含む記述を評価する場合には中間的な評価スケールを追加した3段階評価とする。また、両極端と中間の間にさらに中間的な評価の場合には5段階評価が適切である。各段階評価の評価基準は下記の表3.4による。

表 3.4 段階評価と評価ポイント

評価ポイント	5段階評価	3段階評価	2段階評価
4	明確	N/A	N/A
3	やや明確	N/A	N/A
2	やや不明確	明確	N/A
1	不明確	不明確	明確
0	記載なし	記載なし	記載なし

手順 2 : RFP の形態素解析

ユーザが作成する RFP は様々な形式(例えば, Word 形式, Excel 形式, PDF 形式など)で作成されている. RFP 文章から NFR キーワードを抽出するためには, これらの形式をテキスト形式に変換して処理を行う必要がある. そこで, 自然言語の文章として記述されている部分をテキスト形式に変換する. また, 図として記述されている部分は削除し, 表で記述されている部分は可能な限りテキストに変換する. 以上の処理をモデル構築に用いるすべての RFP について行い, RFP テキストデータ群 R を得る. この RFP テキストデータ群 R をひとつのテキスト文書として形態素解析を行うことにより, RFP テキストデータ群 R の全体から名詞および複合語(名詞語句の組み合わせ)を抽出する(例えば, 「操作」と「手順」の複合語である「操作手順」を抽出する).

手順 3 : NFR キーワードの抽出

テキストマイニングにより抽出された名詞および複合語の中から「小項目」に関係づけられる NFR キーワードを手作業で選択する. この場合, 出現頻度が2回以上の NFR キーワードについて採用し, 1回のみ出現する語句は不採用とするが, 対象とした RFP 群での出現頻度は1回であっても, ソフトウェアアーキテクチャを構築する上で重要と考えられる「小項目」の評価に重要な NFR キーワードは採用した. また, 「小項目」に関する単一の名詞だけでなく専門用語を加えた NFR キーワードにより, より明確に非機能要件の記述内容を評価することが可能となることから複合語を採用した.

複合語として「小項目」の内容を表す助詞を含む複合語句(例えば, 操作を容易にする, 平均読み出し遅延, 障害切り分けなど)は平易な語句で表現された「小項目」を特徴付ける **NFR** キーワードと考えられるので採用した. 複合語の選択にはエキスパートの知識が必要となるが, 本論文で抽出したキーワード群を **Web** に公開している[16]. 一方, 出現頻度は高いが, 広範囲の意味を持つ単語(オペレーションなど)は抽象度が高く, 専門性が低いため「小項目」の記述の明確さを評価するには劣る語句と考えられるので除外する.

次に, 抽出した **NFR** キーワードと「小項目」とのマッピングを行う. k 番目の「小項目」に対応する j 番目の **NFR** キーワードを $W_{kj}(k=1,2,\dots,m, j=1,2,\dots,n)$ と記す. この一連の手作業により, 「**NFR** キーワード評価シート」が得られる. 表 3.5 に「**NFR** キーワード評価シート」の一部を示す. (詳細については, 付録 A. を参照)この表は, 「大項目」, 「中項目」, 「小項目」および「**NFR** キーワード」の 4 層で構成され, 構造化している. 一旦作成された「**NFR** キーワード評価シート」は新たな **RFP** を人手により評価を行う場合においても役立つものとなる.

表 3.5 NFR キーワード評価表（抜粋）

大項目	中項目	小項目	NFRキーワード	
システム運用の評価要件	操作容易特性	オペレーション	操作性, 操作マニュアル, 操作方法, 画面遷移 入力ミス, 操作説明, 容易に操作, 操作手順書 オンラインマニュアル, 直感的に操作	
		応答性	操作説明書, 操作が容易, 操作ガイド, 操作説明 平均応答, ハードディスク応答性能, 秒以下 ネットワーク転送容量, 転送応答性, 最小応答 安定的レスポンス, スループット, データ量 最大スループット, 平均読み出し遅延 ハードディスク容量, レスポンスタイム	
	システム性能	システム性能	ネットワーク使用率, MPU使用率, CPU使用率 ディスクIO負荷率, アクセス速度, 演算性能 ソフトウェア性能, ハードウェア性能, 同時接続数 システム性能, 総合演算性能, 以上の性能	
		負荷バランス	負荷監視, CPU負荷, 回線負荷, 負荷計測 最大負荷時, 負荷分散性能, 負荷分散 ロードバランシング, 負荷率, 負荷低減, ピーク時 負荷低減, ネットワーク負荷, 負荷予測	
	稼働品質	稼働品質	平均稼働率, アクセス量, アクセス頻度 稼働率, %以上, システム稼働率, 正常稼働 システム稼働率, 安定稼働, 正常に稼働 24時間365日稼働, ダウンタイム, 稼働安定性	
	運用監視要件	障害検知特性	システム異常検知	事故発生, 事故発生, 停電信号 イベント発生, 上限値警告, 状態監視 障害通知機能, 状態監視機能, 異常 障害監視機能, 自動運転監視, システム監視
			障害要因	システム障害1次切り分け, 障害発生場所 プラットフォーム障害, 障害発生時刻 障害の切り分け, サーバ障害, ネットワーク障害 ハード障害発生, ネットワーク障害, 電源障害 機器障害, ソフトウェア障害, ハードウェア障害

手順 4 : 評価モデルの構築

「NFR キーワード評価シート」の各「小項目」にマッピングされた NFR キーワード W_{kj} のそれぞれの出現頻度を算出する. NFR キーワードの出現頻度は, 評価する「小項目」ごとに算出する. よって, 評価モデルとなるデータは, 評価対象の RFP の行成分と関連する NFR キーワード W_{kj} の出現頻度である列成分で構成される, 「小項目」ごとのマトリクスとなる. これらのマトリクスが従属変数となり, このマトリクスに手順 1 で作成した手動評価結果を目的変数である教師信号として追加して, ランダムフォレスト推定のための評価モデルを作成する.

3.4.3 「NFR キーワード評価シート」の利用

一般的な RFP 評価として、「NFR キーワード評価シート」を利用するためには、できる限り「小項目」の記述に偏りが無い、さまざまなドメインについての情報化システム RFP 群をサンプルデータとして採用することが求められる。これにより、比較評価対象となる RFP の非機能要件の記述内容の明確さについて、隔たりのない定量評価を行うことが可能となる。また、特定のドメインについて「NFR キーワード評価シート」モデルを利用する場合には、対象とするドメインの RFP 群について、より多くのサンプルデータを採用することが求められる。「NFR キーワード評価シート」モデルは、「小項目」の評価を行うことにより、その上位階層である中項目及び大項目も定量的に評価することができ、階層的に RFP の記述内容の明確さについて確認することができる。記述内容の明確さが不足しているケースでは、記述内容を改良した後に再び「NFR キーワード評価シート」モデルを利用することにより、どの程度非機能要件の記述内容が改良されたかという点について定量的に把握するが可能である。さらに、一度、モデルを構築すれば RFP 作成と評価に多くのエキスパートを必要とする作業を軽減し、少ないコストで非機能要件の記述内容の明確さについて評価を行うことが容易となる。次節では、「NFR キーワード評価シート」モデルにもとづく評価実験について述べる。

3.5 ランダムフォレストを用いた非機能要件評価実験

3.5.1 実験の目的

実験では NFR キーワードに基づく「小項目」の自動評価と手動評価を比較し、教師あり機械学習による自動評価モデルの妥当性を評価する。そこで、5 段階評価、3 段階評価および 2 段階評価の 3 種類の段階評価の評価実験を行い、RFP の多段階評価として、どの評価法が有用であるかについて比較検証する。

3.5.2 評価実験データと実験方法

(1) 評価実験データの準備

評価対象とした RFP サンプルデータは、Web 上に公開されている図書情報システム(11 件), 病院情報システム(10 件), 大学情報システム(8 件), 政府機関情報システム(14 件), 自治体基幹情報システム(10 件), 地方自治体業務システム(14 件)及びその他情報システム(3 件)の計 70 件である. これらの RFP サンプルデータの非機能要件に関する記述部分としてのデータのサイズは, 37,281 行の 1,357,655 語のテキストファイルである.

(2) 教師データの作成

ランダムフォレスト推定の目的変数として与える教師データは 70 件の RFP サンプルデータについて, 「小項目」ごとに評価者が RFP に記述されている内容を評価して段階評価を行う. 評価ポイントは 3.4.2 の手順1で述べた基準に基づいて, 26 個の「小項目」について 2 段階評価, 3 段階評価及び 5 段階評価を行う.

(3) 評価モデル用テキストデータの前処理

評価モデル用のテキストデータをテキストマイニングするツールとして, フリーソフトである KH-Coder[18]を採用する. 評価モデル用の各テキストデータをファイル結合し, HTML タグによって区別された「文書」とみなす単位の集合として構成する. 結合された評価モデル用の RFP 群 R は全体が $\langle h1 \rangle \langle /h1 \rangle$ によりマーキングされ, 各 RFP (r_i) は $\langle h2 \rangle \langle /h2 \rangle$ としてマーキングされる. マーキングした評価モデル用テキストデータの例を以下に示す.

$\langle h1 \rangle$ RFP 非機能要件評価用テキストデータ $\langle /h1 \rangle$

$\langle h2 \rangle$ file:1_A 図書館システム.txt $\langle /h2 \rangle$

RFP 本文.....

.....

<h2>file:15_B 図書館システム.txt</h2>

RFP 本文.....

.....

.....

<h2>file:20_A 病院情報システム.txt</h2>

以下略

テキストマイニングツールでは、RFP 本文内の段落や文を「文書」とみなす単位として HTML タグ<h1></h1>から<h5></h5>までの5つの構造を使用することができるが、本論文では<h1></h1>及び<h2></h2>のみを使用する。これらの前処理の結果、対象としたサンプル RFP の「小項目」を含む評価モデル用のテキストデータが得られる。RFP テキストデータの総語数は 670,125 語である。また、各 RFP についてみると、最大語数、最小語数および平均語数はそれぞれ 79,816 語、1,153 語および 9,573 語である。

(4) キーワードの抽出

最初に、評価モデル用のテキストデータをテキストマイニングツールにより形態素解析を行い、名詞および複合語を対象として NFR キーワードをマニュアルでフィルタリングする。

次に、フィルタリングした NFR キーワードを「小項目」にマッチングする。さらに、マッチングした NFR キーワードを指定して、「小項目」ごとに評価モデル用のテキストデータについてテキストマイニングを行い、26 個の「小項目」に関する NFR キーワードの出現頻度を計測する。この計測結果をもとに、70 個のプロジェクトを「行」とし、各 NFR キーワードの出現回数を「列」とするマトリクスを得る。抽出された名詞および複合語から選択を経て得られた NFR キーワード総数は、776 語である。

(5) 評価実験データの作成

(4) で作成したマトリクスに (2) で作成した 2 段階、3 段階及び 5 段階のマニユ

アル評価結果を対応させて実験用データセットを作成する。評価実験を行うデータセットは、3.4.2 の手順4で述べた手順により作成された「小項目」ごとの 26 個のマトリクスに3種類の段階評価である教師データを付加した計 72 個のデータセットとなる。

(6) 機械学習の条件作成と実験方法

機械学習プログラムは、「R 言語」のランダムフォレストパッケージを採用する。ランダムフォレスト用プログラムは、データ全体の 2/3 を学習データとしてモデルが構築され、構築したモデルを評価するために約 1/3 の残りのデータを取り除いておいてテストデータとする。また、ランダムフォレストの実行にあたっては、学習データと予測データを予測繰り返しの都度ランダムに生成し、ランダムフォレストの分類問題として評価推定を行う。ランダムフォレストの決定木の数はデフォルト値の 500 とする。

実験データは「小項目」ごとに生成しているので、実験回数は 3 種類(2 段階評価, 3 段階評価, 及び 5 段階評価)を評価対象として、26 個の「小項目」について行い、計 72 回のランダムフォレスト推定を実行する。

3.5.3 実験結果の評価手法

ランダムフォレスト推定結果とマニュアル評価の一致率を検定する手法として κ 統計量[31]による検定を行う。 κ 統計量は、ふたりの評定者の判定結果がどの程度一致しているかを示す指標であり、ふたりの評定者の判定が偶然一致する確率を求めて、偶然によらない一致率を以下の計算式により計算する。

$$\kappa \text{ 統計量} = (\text{実際的一致率} - \text{偶然的一致率}) / (1 - \text{偶然的一致率})$$

2段階評価を例にとると、手動評価とランダムフォレストの評価についてのクロス表は以下となる。

		RFによる評価		計
		評価1	評価0	
人手による 評価	評価1	m_{11}	m_{10}	$m_{1\cdot}$
	評価0	m_{01}	m_{00}	$m_{0\cdot}$
計		$m_{\cdot 1}$	$m_{\cdot 0}$	n

ここで、 m_{11} および m_{00} は手動評価とランダムフォレストの評価が一致した RFP の数であり、 m_{01} および m_{10} は不一致となった RFP の数である。実際の一致率 P_e は、 $P_o = (m_{11} + m_{00})/n$ で計算され、偶然の一致率 P_e は、 $P_e = ((m_{11} + m_{10})(m_{11} + m_{01}) + (m_{01} + m_{00})(m_{10} + m_{00}))/n^2$ で計算される。よって、 κ 統計量は下記の式で求められる。

$$\kappa \text{ 統計量} = (P_o - P_e) / (1 - P_e)$$

また、 κ 統計量は、ふたりの評者の判定結果による不一致の程度が多段階評価における評価結果で異なる(例えば、評定者 A が 1 と評定し評定者 B が 2 と評定した場合と、評定者 A が 1 と評定し評定者 B が 3 と評定した場合)場合に、より不一致であるとする結果に対して重みを大きくする「重み付け κ 統計量」による検定がある。

本実験では、評価者による評価結果とランダムフォレスト推定評価結果により不一致がある場合には重みを付けることが適当と考え、「重み付け κ 統計量」(以後単に κ 統計量とする)により検定する。また、 κ 統計量による評価を行った結果から、3 種類の段階評価の中でどの段階評価が最も有効であるかについて考察する。以下は κ 統計量による判定基準である。

κ 統計量による判定基準

- 0 ~ 0.40 : 低い一致
- 0.41 ~ 0.60 : 中等度の一致
- 0.61 ~ 0.80 : かなりの一致
- 0.81 ~ : 高い一致

さらに、「3 段階評価」については、ランダムフォレストによるテストデータの評価推定結果が評価者の評価と一致しているかどうかを下記の一一致判定基準についても考察する。

一致率 = 一致ケース / 全テストデータケース

±1 差一致率 = ((一致ケース) + (±1 差一致ケース)) / 全テストデータケース

ここで、

一致ケース: データセットをランダムサンプリングして 1/3 をテストデータとした場合にランダムフォレストによる予測と教師データが一致したケース数とする。

全テストデータケース: 70 件のデータセットをランダムサンプリングした 1/3 のテストデータ数とする。

±1 差一致ケース: データセットをランダムサンプリングして 1/3 をテストデータとした場合にランダムフォレストによる予測と教師データが一致した場合とその評価の差が ±1 差であったケースの合計数とする。

3.5.4 実験結果に基づく一致の検定と考察

3 段階評価について算出した「小項目」の κ 統計量による検定と一致率の結果を、図 3.7 にレーダチャートで示す。図 3.7 によると、 κ 統計量は一致率と比較して各「小項目」の評価に差が大きく表れることが分かる。そこで、より階層の高い「中項目」の評価を κ 統計量により比較して、平均化された κ 統計量をもとに 3 種類の段階評価の内どの段階評価の一致率が高いかについて考察する。「中項目」の一致率は、各「中項目」に属するすべての「小項目」について算出した κ 統計量の平均値としている。「中項目」についての 3 種類の段階評価の κ 統計量を表 3.6 示す。表 3.6 によると、2 段階評価の κ 統計量 0.3591 は、これ以外の多段階評価と比較して、「冗長化特性」と「セキュリティ対策特性」は比較的高い一致と言えるが、高い一致(稼働品質特性, データバックアップ特性, 冗長化特性など)と低い一致(障害検知, 障害予防, 障害復旧など)の差が大きく、安定していないことが認められる。5 段階評価も同様に高い一致と低

い一致率の差が大きく、安定しているとは言えない。3段階評価の「中項目」全体の κ 統計量は0.4063であり、2段階評価と5段階評価と比較して個々の「中項目」について的一致差も少なく、全体の κ 統計量として最も高い数値を示している。さらに、3段階評価の個々の「中項目」で見ると、最も高い一致を示した「システム監視」(κ 統計量が0.6487)と最も低い一致を示した「障害復旧」(κ 統計量が0.2026)があり、それぞれの「中項目」で大きい差があることが認められる。これは、 κ 統計量の計算において一致の隔たり(クロス表の対角部分の隔たり)があれば一致が低くなり、 κ 統計量が小さくなることが要因と考えられる。そこで、例として「システム監視」の小項目である「運用管理」および「システム管理」と「障害復旧」の「小項目」である「リカバリ処理」および「パッチ処理」のクロス表と κ 統計量を図3.8に示す。図3.8によるとクロス表で一致の隔たりがある「リカバリ処理」および「パッチ処理」の κ 統計量が低いことが認められる。

よって、「小項目」の個々の一致については、実験結果による κ 統計量に差はあるが、3種類の段階評価の内、3段階評価が最も高い一致を示すことが分かった。次項3.5.5では、ランダムフォレストによる3段階評価と5段階評価についての自動評価と手動評価の結果を比較検証する。

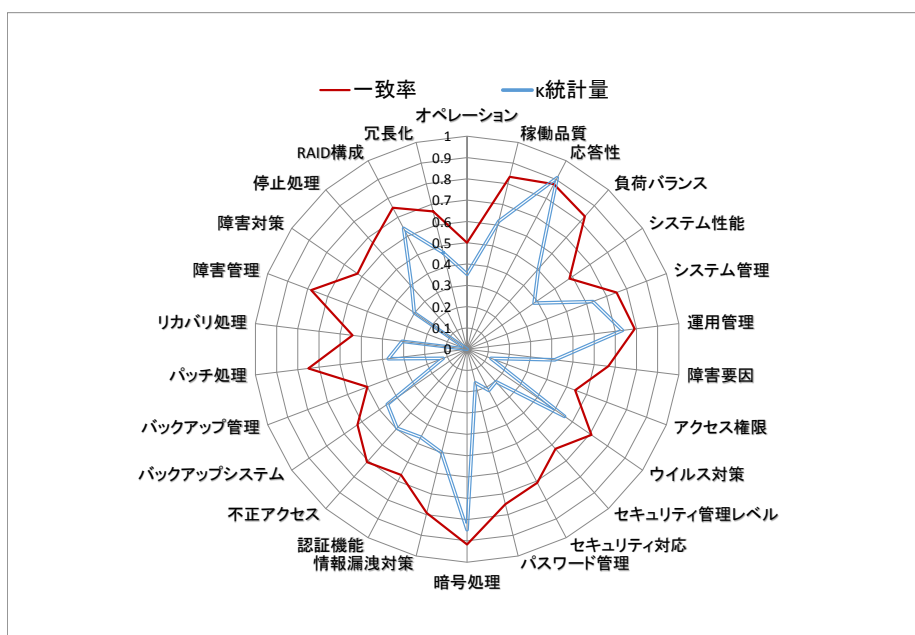


図 3.7 「小項目」の3段階評価の比較

リカバリ処理					
$\kappa =$	RFによる評価				計
	評価2	評価1	評価0		
0.2931					
手動評価	評価2	0	0	0	0
	評価1	2	4	0	6
	評価0	1	8	8	17
	計	3	12	8	23

パッチ処理					
$\kappa =$	RFによる評価				計
	評価2	評価1	評価0		
0.2780					
手動評価	評価2	0	0	0	0
	評価1	0	2	0	2
	評価0	1	4	16	21
	計	1	6	16	23

運用管理					
$\kappa =$	RFによる評価				計
	評価2	評価1	評価0		
0.7333					
手動評価	評価2	1	1	0	2
	評価1	5	11	0	16
	評価0	0	0	6	6
	計	6	12	6	24

システム管理					
$\kappa =$	RFによる評価				計
	評価2	評価1	評価0		
0.6308					
手動評価	評価2	0	0	0	0
	評価1	4	11	0	15
	評価0	0	2	7	9
	計	4	13	7	24

図3.8 クロス表と κ 統計量（一部）

表 3.6 段階評価方法別の κ 統計量

非機能要件 中項目	2段階評価 κ 統計量	3段階評価 κ 統計量	5段階評価 κ 統計量
オペレーション	0.2727273	0.352601	0.2814371
稼働品質	0.538645	0.632996	0.5681375
障害検知	0.225141	0.416683	0.18433735
システム監視	0.343091	0.648718	0.467636
セキュリティ対策	0.472326	0.414079	0.294958
冗長化	0.628205	0.486667	0.4251353
データバックアップ	0.51875	0.268439	0.3625605
障害予防	0	0.234274	0.10996563
障害復旧	0.2333333	0.20265	0.07539316
平均	0.3591	0.4063	0.3077

3.5.5 手動評価と自動評価の比較

ランダムフォレストによる 1/3 のテストデータの推定評価結果について、手動評価との一致率および ± 1 差一致率を計算(3.5.3 項参照)した 3 段階評価の結果を表 3.7(a)に示す。また、5 段階評価については表 3.7(b)に示す。

3 段階評価による 26 個の「小項目」について的一致率及び ± 1 差の一致率の平均値は、それぞれ、69.8%及び 97.2%で高い一致であった。また、一致率が 70%を超えたものは「小項目」26 個の内の 13 個であり、全体の約 1/2 であった。更に、 ± 1 差一致率では、25 個が 90%を超えている。このことは、評価が

全く異なる評価点の差が“2”である RFP が少ないことによる。ただし、すべての RFP について「小項目」26 個の評価点が“3”と評価された高評価の RFP は、手動評価全体の 9.8%とその比率が少ない。一方、 κ 統計量の平均は 0.43 であり、ランダムフォレストによる推定評価と評価者による評価との正答率は約 43%であった。「小特性」ごとに κ 統計量の一致基準で見ると、「小項目」26 個の内、「高い一致」が 2 個（「応答性」および「暗号処理」）、「かなりの一致」が 4 個（「稼働品質」、「運用管理」、「システム管理」および「RAID 構成」）および「中程度の一致」が 8 個であり、「中程度の一致」以上と判定される「小特性」の比率は 54%である。これは、RFP サンプルデータの非機能要件の記述内容が、明確に書かれているものが少なく、評価点が“1”および“2”のデータが多く、クロス表の対角部分の隔たりが大きいことが要因のひとつと考えられる。この結果、 κ 統計量によってランダムフォレストの推定結果と手動評価との一致を評価すると低い一致となることが考えられる。よって、「小項目」の記述内容の明確さが高いプロジェクトの RFP をサンプルデータとして収集し、評価点が“3”の高い評価の RFP サンプルを含めた学習データによる自動評価を行うことにより κ 統計量による一致の評価も高くなることが期待される。

一方、5 段階評価の一致率は 61.96%であり、3 段階評価の一致率である 69.84%との差は 7.88%となり、一致率が低いと言える。また、 ± 1 差一致率の場合は 12.11%となり、その差はさらに大きいことが認められる。3 段階評価の ± 1 差一致率が 5 段階評価より高くなっている要因として、多段階評価における 3 段階評価の中心性向が高いことが考えられる。さらに、5 段階評価では、ランダムフォレストの分類精度（3 段階評価の平均 OOB 誤差は 28.91%であり 5 段階評価の平均 OOB 誤差は 38.71%）が低いために一致率が低くなっているものと考えられる。

次に、3 段階評価の場合のランダムフォレストによる「自動評価」と「手動評価」平均値について、誤差および有意水準 5%で t 検定を行った結果を表 3.8 に示す。表 3.8 によると「小項目」は 26 個の内 15 個に有意差があり、評価が一致していると判定される「小項目」は全体の約 42%の 11 個であった。これらの結果から、ランダムフォレストによる「自動評価」においては、3 段階評価が 5 段階

評価より「手動評価」と一致していると言える。

さらに、各「小項目」のランダムフォレストによる 1/3 のテストデータについての「自動評価」の平均値と対応する「手動評価」の平均値を図 3.9(a)および 3.9(b)のレーダチャートに示す。段階評価による評価結果をわかりやすくするために、評価点を“1”，“2”，および“3”の 3 段階と“1”，“2”，“3”，“4”および“5”の 5 段階評価としている。図 3.9 によると、「応答性」，「運用管理」，「アクセス権限」及び「バックアップ管理」については、ランダムフォレストモデル推定評価と評価者による評価がほぼ一致しているが、その他の「小項目」ではいずれも評価者による評価が高い結果となっていることが認められる。さらに、5 段階評価と 3 段階評価との比較においては、5 段階評価は「自動評価」と「手動評価」の評価差がより大きいことが認められる。

表 3.7(a) ランダムフォレストの評価推定結果(3 段階評価)

小項目	一致数	±1差 一致数	不一致 数	一致率	±1差 一致率	κ 統計量
オペレーション	12	11	1	50.00%	95.83%	0.353
稼働品質	20	2	2	83.33%	91.67%	0.619
応答性	21	3	0	87.50%	100.00%	0.909
負荷バランス	20	3	1	83.33%	95.83%	0.505
システム性能	14	10	0	58.33%	100.00%	0.381
システム管理	18	6	0	75.00%	100.00%	0.631
運用管理	19	5	0	79.17%	100.00%	0.733
障害要因	16	7	1	66.67%	95.83%	0.415
アクセス権限	13	11	0	54.17%	100.00%	0.120
ウイルス対策	17	7	0	70.83%	100.00%	0.556
セキュリティ管理レベル	15	7	2	62.50%	91.67%	0.204
セキュリティ対応	17	7	0	70.83%	100.00%	0.219
パスワード管理	18	3	3	75.00%	87.50%	0.163
暗号処理	22	2	0	91.67%	100.00%	0.848
情報漏洩対策	19	5	0	79.17%	100.00%	0.500
認証機能	16	8	0	66.67%	100.00%	0.467
不正アクセス	17	7	0	70.83%	100.00%	0.495
バックアップシステム	15	9	0	62.50%	100.00%	0.455
バックアップ管理	12	11	1	50.00%	95.83%	0.120
パッチ処理	18	5	1	75.00%	95.83%	0.373
リカバリ処理	13	10	1	54.17%	95.83%	0.307
障害管理	18	5	0	78.26%	100.00%	0.000
障害対策	15	9	0	62.50%	100.00%	0.301
停止処理	16	8	0	66.67%	100.00%	0.392
RAID構成	18	6	0	75.00%	100.00%	0.640
冗長化	16	8	0	66.67%	100.00%	0.467
全体	435	175	13	69.84%	97.92%	0.430

表 3.7(b) ランダムフォレストの評価推定結果(5段階評価)

小項目	一致数	±1差 一致数	不一致数	OOB誤差	一致率	±1差 一致率
オペレーション	13	4	6	58.70%	56.52%	73.91%
稼働品質	19	3	1	34.78%	82.61%	95.65%
応答性	23	0	1	0.00%	95.83%	95.83%
負荷バランス	21	0	3	19.57%	87.50%	87.50%
システム性能	12	7	5	47.83%	50.00%	79.17%
システム管理	17	4	3	32.61%	70.83%	87.50%
運用管理	14	5	4	65.22%	60.87%	82.61%
障害要因	10	9	5	52.17%	41.67%	79.17%
アクセス権限	9	10	5	60.87%	37.50%	79.17%
ウイルス対策	14	8	1	32.61%	60.87%	95.65%
セキュリティ管理レベル	11	8	5	50.00%	45.83%	79.17%
セキュリティ対応	10	11	3	50.00%	41.67%	87.50%
パスワード管理	14	4	6	47.83%	58.33%	75.00%
暗号処理	20	2	2	4.35%	83.33%	91.67%
情報漏洩対策	17	7	0	15.22%	70.83%	100.00%
認証機能	15	5	4	50.00%	62.50%	83.33%
不正アクセス	16	6	2	34.78%	66.67%	91.67%
バックアップシステム	9	14	1	52.17%	37.50%	95.83%
パッチ処理	17	3	4	39.13%	70.83%	83.33%
リカバリ処理	10	8	6	41.30%	41.67%	75.00%
障害管理	22	0	2	17.39%	91.67%	91.67%
障害対策	13	9	2	45.65%	54.17%	91.67%
停止処理	14	7	3	34.78%	58.33%	87.50%
RAID構成	16	6	2	32.61%	66.67%	91.67%
冗長化	14	6	4	39.13%	58.33%	83.33%
システム異常検知	14	2	8	47.83%	58.33%	66.67%
全体	384	148	88	38.71%	61.96%	85.81%

表 3.8 自動評価と手動評価の誤差

小項目	RF評価得点	手動評価得点	誤差(手動評価-RF評価)	P(T<=t)両側
オペレーション	1.33	1.71	0.38	0.017
応答性	1.88	1.92	0.04	0.575
システム性能	1.21	1.63	0.42	0.000
負荷バランス	1.17	1.29	0.13	0.266
稼働品質	1.17	1.42	0.25	0.056
障害要因	1.42	1.71	0.29	0.056
運用管理	1.88	2.00	0.13	0.185
システム管理	1.63	1.88	0.25	0.011
不正アクセス	1.38	1.58	0.21	0.057
アクセス権限	2.04	1.92	0.13	0.377
セキュリティ管理レベル	1.17	1.63	0.46	0.002
セキュリティ対応	1.04	1.92	0.88	0.000
認証機能	1.50	1.75	0.25	0.031
パスワード管理	1.04	1.42	0.38	0.017
ウイルス対策	1.38	1.67	0.29	0.005
情報漏洩対策	1.17	1.38	0.21	0.022
暗号処理	1.38	1.46	0.08	0.162
冗長化	1.38	1.54	0.17	0.162
RAID構成	1.33	1.58	0.25	0.011
バックアップ管理	2.00	1.96	0.04	0.802
バックアップシステム	1.50	1.79	0.29	0.016
障害対策	1.38	1.42	0.04	0.747
障害管理	1.00	1.21	0.21	0.022
停止処理	1.17	1.50	0.33	0.003
リカバリ処理	1.25	1.75	0.50	0.000
バッチ処理	1.13	1.42	0.29	0.016
平均	1.38	1.63	0.26	

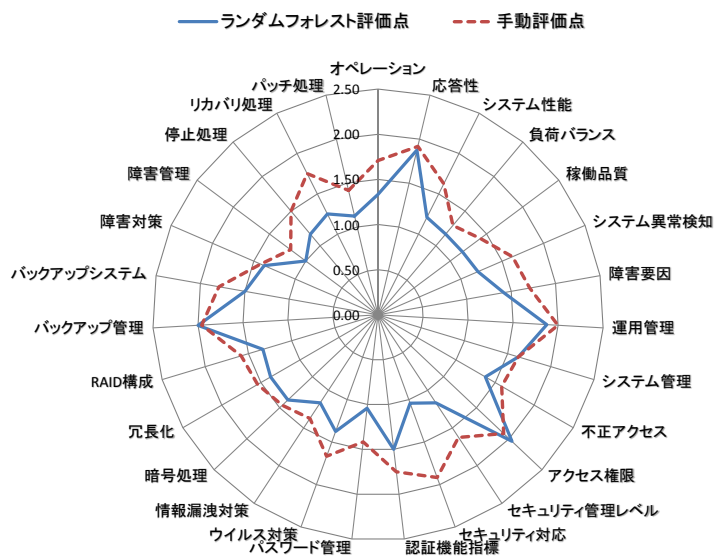


図3.9(a) 自動評価と手動評価の評価得点レーダチャート(3段階評価)

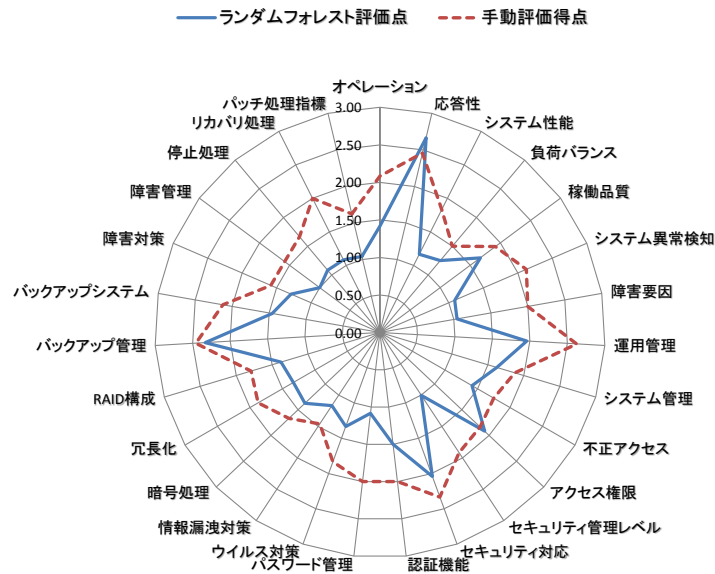


図3.9(b) 自動評価と手動評価の評価得点レーダチャート(5段階評価)

図 3.9 自動評価と手動評価の評価得点レーダチャート

3.6 まとめ

本章では、NFR キーワードを RFP からテキストマイニングにより抽出し、「非機能要件評価シート」に追加した「小項目」とマッピングすることにより、「非機能要件評価シート」の改良を行い、「NFR キーワード評価シート」を作成した。次に、RFP の「小項目」の記述内容の明確さを NFR キーワードにより自動評価することを目的として、「NFR キーワード評価シート」の重要な「小項目」を教師あり機械学習の手法であるランダムフォレストによる評価推定モデルを構築した。ランダムフォレストによる評価結果については、2 段階、3 段階及び 5 段階の多段階評価の中でどの評価手法が有効であるかについて「小項目」ごとに評価を行い、「中項目」を対象に有効性の評価を行った。評価結果の検証については、ランダムフォレストによる自動推定評価と評価者による手動評価を一致率の計算と統計的手法である κ 統計量の計算により一致の考察を行った。

実験の結果、3 段階評価が他の段階評価と比較して高い一致を示した。

3 段階評価について、ランダムフォレストによる推定結果と評価者による評価が

一致しているかどうかについて計算した一致率は、69.8%であった。ただし、一致率が70%を超えたものは「小項目」26個の中の13個であり、「小項目」についての評価が高い評価対象としたRFPが少ないという問題があった。一方、 κ 統計量の判定基準では「中程度の一致」以上が54%であった。これらの実験結果は、NFRキーワードによるRFPの非機能要件の記述内容の明確さを評価することは可能であり有用であることを示している。ただし、「小項目」の評価が高いRFPをさらに追加して、評価対象のRFPの評価結果の隔たりを減少させることが今後の課題である。

次に、自動評価モデルの誤差について評価を行った結果、自動評価と手動評価が一致すると判断される「小項目」は26個の中の11個であり、全体の約42%であった。これらの結果から、「小項目」をNFRキーワードで自動評価するモデルは有効であると言えるが、NFRキーワードの選定や数を変えてモデルのチューニングを行い、ランダムフォレストの精度を高めることが今後の課題である。

本章で提案した自動評価モデルは、NFRキーワードによるRFPの非機能要件を評価することの可能性を示すものであり、RFPに記述された非機能要件の記述が明確であるかどうかを自動評価した結果を利用することにより、ユーザとベンダの技術討論を通じて委託ソフトウェア開発プロセスの上流工程でRFPの品質向上に貢献できる有効なモデルであると考えられる。

第4章 機械学習によらない非機能要件の自動評価

4.1 概要

第3章では、「NFR キーワード評価シート」を作成し、NFR キーワードによる「小項目」の自動評価モデルを提案[43]し、RFP に記述された非機能要件の明確さの評価について有効性を確認した。しかし、教師あり機械学習による自動評価モデルでは、学習データの隔たりおよび RFP サンプル数が多い場合の教師データ作成にかかる工数について課題があった。

そこで、本章ではウェブから収集した 161 件の RFP サンプルデータ群より抽出した NFR キーワードを「小項目」に関連付け、RFP の隔たりを減少させた評価データを対象とした。（評価対象とした「小項目」と関連づけした NFR キーワードについては付録 B を参照のこと）そして、RFP サンプル数を多くした場合でも、RFP に記述された「小項目」を教師なしで評価する機械学習によらない 3 つの重み付け手法について提案し、比較評価する。そこで、3 つの手法の比較を行うために、161 件の RFP について手動評価を行い、手動評価との比較により有効な手法を検証する。また、追加実験として教師なし機械学習の一つであるクラスタリングによる評価実験を行う。3 つの手法は、シンプル TF-IDF 法によるもの、TF-IDF 法の考えを拡張したものおよびトークン比の考え方を応用したキーワード比によるものである。それぞれの手法では、NFR キーワードの出現頻度による重みを計算した「NFR キーワード指数」により評価スコアを算出して評価する。以下、4.2 節で関連研究について述べる。4.3 節では提案手法の詳細を述べ、4.4 節では実験結果による 3 つの手法の比較評価を行う。さらに、4.5 節では拡張 TF-IDF とキーワード比を用いた RFP 評価のクラスタリングについての追加実験について述べ、4.6 節でまとめを述べる。

4.2 関連研究

機械学習によらないで、自然言語で記述された非機能要求内容の品質を評

価する先行研究は、(1)自然言語で記述された要求の品質を構文解析あるいは言語としての欠陥評価をするモデルに関する研究[34] [35]、(2)オントロジを用いた要求抽出と分析[36] [41] [44]、(3)コーパスに基づく要求分析[37]、(4)パターンモデルにもとづく分類による要求分析モデルの研究[26] [28]などがある。しかし、(1)はソフトウェア開発段階でソフトウェア要求仕様書の評価をエンジニアが行うためのツールであり、本研究が対象としている超上流工程でのユーザ要求の評価モデルとしては適当ではない。(2)はオントロジを用いて欠落や矛盾のないソフトウェア仕様書を作成することを目的としており、要求品質を評価してソフトウェア設計に反映させることに有効であるが、評価に時間を要する。(3)は要求文書を分析することを目的としており、要求品質の定量的評価に言及していない。(4)は抽出語句を **NFR** 分類器により分類し、**NFR** に関連付けを自動的に行うが、プロジェクトの定量評価は行っていない。

また、IPA SEC では、「ドキュメント品質評価指標」として、**NFR** の記述十分性を要求仕様書内の各パートのページ数と要求仕様書ページ数の総和の比を用いた“非機能要件記述バランス”を提案している[45]。

ドキュメント品質評価指標：要求仕様書バランス

= (要求仕様書内の各パートのページ数) / (要求仕様書ページ数の総和)

これを **NFR** キーワードで置き換えて表現すると

非機能要件記述十分性評価指標：**NFR** キーワード記述バランス

= (各 **RFP** の **NFR** キーワード出現回数の合計) / (**RFP** テキスト全体の **NFR** キーワード出現回数)

となるが、個別の **NFR** 記述内容について評価することはできない。

また、「非機能要求グレード評価」による非機能要求グレード抽出項目は筆者らによる先行研究の「非機能要件評価シート」に該当し、非機能要件を決定するプロセスを支援するが、非機能要件の記述内容について具体的なレベルでの定量評価は行っていない。また、**RFP** から非機能要求グレードを抽出するプロセスに多くの時間をかける必要がある。

本章で提案する評価モデルは、「小項目」ごとに自動的に定量的評価することが可能であり、より具体的な **NFR** について記述内容をさらに明確にするべき

かどうかを定量的に把握することが可能である点において優位性がある。

4.3 提案手法

本節では、161 件の RFP から抽出した NFR キーワードの出現頻度を用いて算出する 3 つの提案手法の NFR キーワードの重み付けを示し、NFR 評価スコアの算出方法について述べる。

4.3.1 NFR キーワードの抽出

収集した 161 件の RFP テキストを結合し、各 RFP を文書単位としたテキストデータを対象として、テキストマイニングにより NFR キーワードを抽出する。抽出した NFR キーワードは、「小項目」の内、第 3 章で述べたランダムフォレストの予測結果とマニュアル評価の結果が「低い一致」であった「負荷バランス」、「アクセス管理」、「セキュリティ管理」、「障害対策」および「リカバリ」と「高い一致」を示した「応答性」について見直しを行い、マッピングした。NFR キーワードの追加に当たっては、出現頻度が少ない場合でも、長くて難しい複合語であれば意味の重要性が高いとする「語彙の意味重要性」(たとえば、「ロードバランシング構成」、「情報セキュリティマネジメントシステム」、「情報システムダウン対策マニュアル」など)の考えに基づいて、NFR に関する複合語を追加した。次項で NFR キーワードへの重み付け手法である「NFR キーワード指数」について述べる。

4.3.2 NFR キーワードの重み

「小項目」にマッピングされた多種の NFR キーワードについての出現頻度が高い RFP は、「小項目」の記述内容の明確さについて評価が高いことが推定される。しかし、特定の NFR キーワードが同一の RFP に集中して出現する場合は、その出現の偏りが大きく、重要な NFR キーワードとは限らない。また、出現頻度が少ない NFR キーワードであっても、長く難しい複合語としての専門用語

であれば重要性があり、「小項目」の記述内容の明確さが高いことが考えられる。そこで、対象とした RFP の中で、個々の NFR キーワードがどの程度重要であるかを示す指数である「NFR キーワード指数」を「小項目」にマッピングされた NFR キーワードの出現頻度から(算出方法は 4.3.3 項を参照)求めることが必要となる。

一般的に、文章における語彙の重要度を重みとして数量的に示す方法として、TF-IDF 法が広く用いられている。TF-IDF 法は経験的な方法であり、さまざまな拡張が提案されている。TF-IDF 法は文章から特徴的なキーワードを抽出し、文章に記述されている内容を定量的に特徴づける手法として適用されるケースが多い。しかし、TF-IDF 法の問題点として必ずしも文章を特徴づける語句が高いスコアになるとは限らない。

次項では、3 つの手法である単純 TF-IDF、拡張 TF-IDF およびキーワード比の定義と、これらにより算出した重みとしての「NFR キーワード指数」について述べる。

4.3.3 単純及び拡張 TF-IDF とキーワード比

最も一般的な TF-IDF 法は文書 d 内の単語 w に以下の重みを与えると定義されている[32]。

$$\text{TF-IDF_Weight}(w,d) = \text{TermFreq}(w,d) * \log(N/\text{DocFreq}(w))$$

ここで、 $\text{TermFreq}(w,d)$ は文書 d 内での単語 w の出現頻度、 N は全文書数、 $\text{DocFreq}(w)$ は単語 w を含む文書数である。

TF は文書 d における語 w の頻度であり、頻度が高いほど大きな値をとる。一方、IDF は語 w が多くの文書に表れるほど小さくなり、特定の文書に表れるほど大きくなる。

しかし、シンプルな TF-IDF 手法では NFR キーワードの出現頻度が重みとして極めて大きく影響する。提案手法は、複数の文章の集合である RFP に対して「小項目」による個別評価を目的としており、同一の NFR キーワードが特定の RFP に集中して出現する場合の極端な重み付けによる評価差を避ける必要がある。また、多

数の RFP に出現する NFR キーワードは重要性が高いと考えられる。そこで、2 番目の手法として各「小項目」にマッピングした NFR キーワードの極端な重み付をなくすための手法としてシンプルな TF-IDF を拡張した手法を提案している。

TF-IDF の拡張については様々な方法が提案されているが、「Log および文書中のタームにより正規化する」方法を採用し、下記の式により拡張 TF-IDF である Ex_TF-IDF_Weight(w_{sub}, d_{rfp})を算出する。対数の底は2としている。この手法により、出現頻度が極めて高い NFR キーワードの評価を出現頻度が極めて低い NFR キーワードに対してその差を平準化することができる。

$$\text{Ex_TF-IDF_Weight}(w_{sub}, d_{rfp}) = \text{Ex_TermFreq}(w_{sub}, d_{rfp}) * \text{Ex_IDF}(w_{rfp})$$

ここで、

$$\text{Ex_TermFreq}(w_{sub}, d_{rfp}) = \frac{1 + \log_2 w_{sub}}{\log_2 w_{subn}}$$

$$\text{Ex_IDF}(w_{rfp}) = \log_2 (N_{rfp}/w_{rfp})$$

ここで、

Ex_TermFreq : RFP テキスト全体における NFR キーワード出現頻度

w_{sub} : RFP テキスト全体における各「小項目」に属する NFR キーワードの出現頻度

w_{subn} : 各「小項目」に属する NFR キーワード種類の総数

Ex_IDF : RFP テキスト全体における NFR キーワードが出現する RFP 数

N_{rfp} : 評価対象とした RFP の総数(定数:本実験では 161)

w_{rfp} : 各「小項目」に属する NFR キーワードが出現した RFP の総数

3 番目の手法としてトークン比の考え方に基づいた語彙の豊富さによる重み付けについて述べる。ここで、トークン比[39]とは語彙の豊富さを示す最も簡単な指標であり、延べ語数 N に対する異なり語数 V の比率(Type-Token Ratio) $TTR = V/N$ で示される。本手法のキーワード比は以下の式で算出する。

$$TKR = V_k/N_{sub}$$

ここで、

TKR:キーワード比(Type-KeyWord Ratio)

Vk:RFP 中に出現する「小項目」に関する NFR キーワードの種類数

Nsub:「小項目」に属する NFR キーワードの総数

161 件の RFP より抽出し、「応答性」にマッピングされた NFR キーワードの出現頻度により算出した拡張 TF-IDF, 単純 TF-IDF およびキーワード比の 3 つの手法による重みと NFR キーワード出現頻度の一部を表 4.1 に示す. 表 4.1 によると, 単純 TF-IDF やキーワード比については, たとえば「スルーブット」という NFR キーワードの場合のように出現頻度が高い場合には, NFR キーワードの出現頻度の影響が大きく, 重みは極めて大きな値をとるが, 拡張 TF-IDF は相対的に極めて低い値に抑えられることがわかる. このことは, 拡張 TF-IDF の場合には, 多くの RFP に頻繁に出現する NFR キーワードの極端な重み付けによる評価点を低く抑えることがわかる. 一方, 単純 TF-IDF やキーワード比は多くの RFP に頻繁に出現する NFR キーワードの極端な重み付けによる評価差が大きくなる. また, キーワード比については特定の RFP に集中して出現する NFR キーワードの重み付けについては, その他の手法と比較して低く抑えられている.

図 4.1 に単純 TF-IDF と NFR キーワード出現頻度を出現数の多い NFR キーワード順に並べた場合の関係を示し, 図 4.2 に拡張 TF-IDF と NFR キーワードが出現する RFP 数を出現 RFP 数が多い NFR キーワード順に並べた場合の関係を示す. 図 4.1 によると, NFR キーワードの出現頻度が高い場合において単純 TF-IDF は極めて変動が大きいことが認められ, 拡張 TF-IDF は NFR キーワードが出現する RFP 数が中位の領域において相対的に高い値となり, NFR キーワード出現数に大きく影響されないことを示している. このことは, 拡張 TF-IDF による「NFR キーワード指数」が, 特定 RFP に集中して出現する一般性に乏しい NFR キーワードが極端に高い評価ポイントとなることを低減するフィルタとして働き, 比較的多くの RFP に出現するキーワードの重要性を上げる働きを持っていることがわかる. 次項では, 「NFR キーワード指数」による RFP の評価スコア算出の方法について述べる.

表4.1 3つの手法による重みとNFRキーワードの出現頻度（抜粋）

	NFRキーワード	拡張TF-IDF	単純TF-IDF	キーワード比	出現頻度	出現RFP数
応答性	応答性	4.066	47.460	0.182	10	6
	最小応答	2.944	17.238	0.091	3	3
キーワード 語数:33	平均応答	2.510	12.662	0.061	2	2
	最小レスポンス	1.453	7.331	0.030	1	1
	応答速度確保	2.907	14.662	0.030	2	1
	安定的レスポンス	2.907	14.662	0.030	2	1
	転送応答性	2.907	14.662	0.030	2	1
	端末レスポンス	2.510	12.662	0.061	2	2
	ネットワーク応答	1.453	7.331	0.030	1	1
	画面レスポンス	4.169	31.655	0.061	5	2
	ハードディスク応答性能	1.453	7.331	0.030	1	1
	平均読み出し遅延	2.510	12.662	0.061	2	2
	システム応答速度	1.453	7.331	0.030	1	1
	スループット	4.040	257.958	0.758	96	25
	スループット目標値	6.493	149.395	0.091	26	3
	最大スループット	2.510	12.662	0.061	2	2
	VPNスループット	2.510	12.662	0.061	2	2
	応答時間	3.568	129.060	0.727	47	24
	オンライン応答時間	3.560	30.054	0.152	6	5
	レスポンスタイム	3.662	145.536	0.727	53	24
	レスポンス目標値	8.584	219.928	0.030	30	1
	ターンアラウンド	4.687	95.280	0.242	22	8
	秒程度	3.467	68.113	0.515	21	17
	秒以内	2.650	259.685	1.606	162	53
	秒以下	4.315	140.385	0.455	41	15
	タイムラグ	3.511	26.655	0.121	5	4
	平均処理応答	2.907	14.662	0.030	2	1
	フォワーディングレート	4.828	36.655	0.030	5	1
	メモリ使用率	3.440	37.449	0.273	9	9
	主記憶容量	6.081	248.796	0.212	55	7
	ハードディスク容量	4.838	136.306	0.303	34	10
	ネットワーク転送容量	3.757	21.993	0.030	3	1
	データ量	2.292	197.940	1.818	139	60

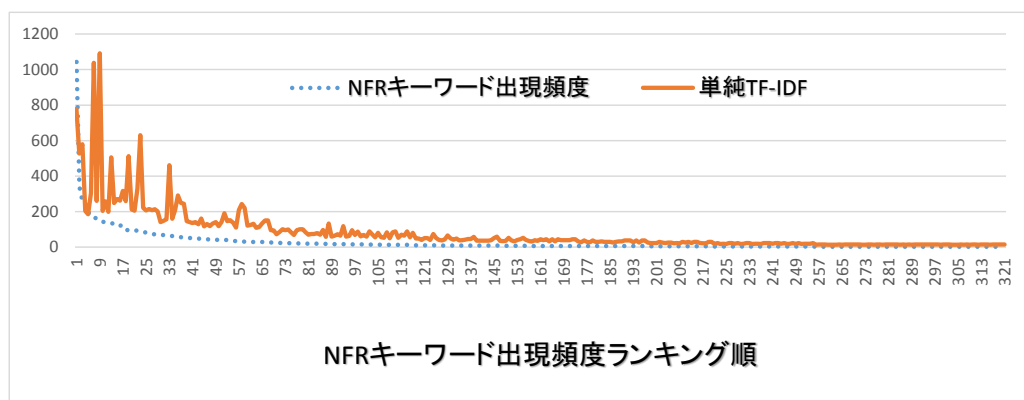


図 4.1 単純 TF-IDF と NFR キーワード出現頻度

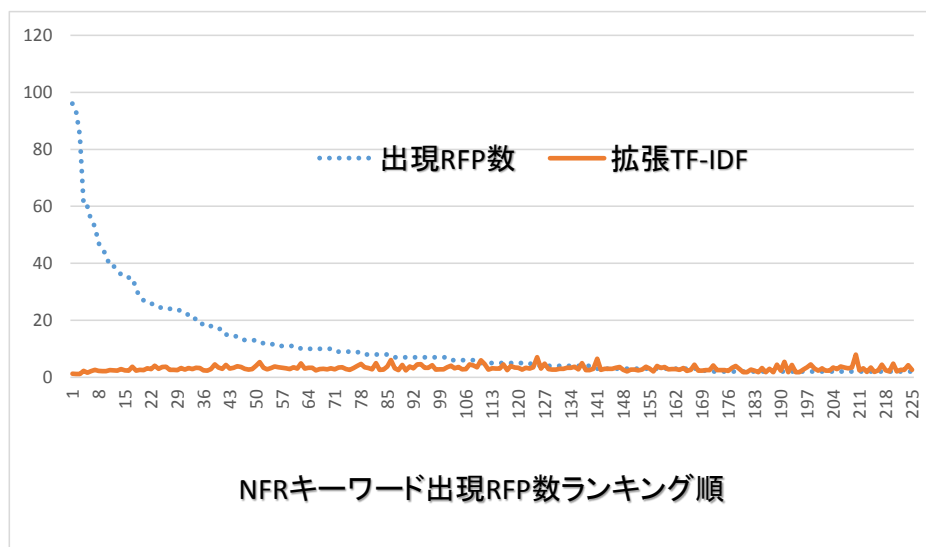


図 4.2 拡張 TF-IDF と NFR キーワード出現 RFP

4.3.4 RFP の NFR 評価スコア

提案する手法のもとになった TF-IDF 手法は、対象とする文書全体において語句が特徴的に多く出現することを評価する指標である。本研究の目的は、多くの RFP に特徴的に表れる NFR キーワードにより非機能要件である「小特性」の評価スコアを求め、個々の RFP を定量的に評価することである。したがって、NFR キーワードの出現頻度をそのまま用いると、一つの RFP に繰り返し記述され出現する NFR キーワードにより、その RFP の評価スコアが高い結果となる。しかし、一つの NFR キーワードの出現頻度が高い場合に、その RFP の非機能要件の記述内容が明確であるとは限らない。そこで、個々の RFP を文書単位として、評価対象としたすべての RFP から NFR キーワード抽出し「小項目」にマッピングする。そして、各手法により算出した「NFR キーワード指数」による重み付けを行い、「小項目」ごとに RFP の評価スコアを算出している。こうすることにより、特定の RFP のみに集中的に出現した NFR キーワードがこの特定の RFP の評価スコアを高くすることを避けることが可能となる。

「小項目」ごとの評価スコア算出に当たっては、最初に、テキストマイニングを行い、各 NFR キーワードの出現頻度を列とし、各 RFP を行としたマトリクスを生成する。次

に、このマトリクスにおいて各 **NFR** キーワードが出現したかどうかについて **NFR** キーワードが該当 **RFP** に出現した場合を“1”とし、出現しなかった場合を“0”としたマトリクスに変換する。したがって、**RFP** ごとの **NFR** 評価スコアは、これらの重みを合計して算出される。以下に **NFR** 評価スコア算出式を示す。

$$RFPscore(i) = \sum RFP(i)[1,1,0, \dots, 0,1,0] * KW(i, j)$$

ここで、

$RFPscore(i)$: 「小項目」 i の **NFR** 評価スコア

$RFP(i)[1,1,0, \dots, 0,1,0]$: 「小項目」 i の **NFR** キーワード出現行列

$KW(i, j)$: 「小項目」指標 i に属する j の「**NFR** キーワード指数」

表 4.2 に拡張 **TF-IDF** により算出した「小項目」ごとの **NFR** キーワード数、及び「**NFR** キーワード重み」の例を示す。表 4.2 によると拡張 **TF-IDF** の場合に、「小項目」にマッピングされた **NFR** キーワード種類にかかわらず **NFR** キーワードの重み付けされていることが分かる。

表4.2 NFRキーワード重みの例

小項目	NFRキーワードの例	NFRキーワード重み
応答性	応答性	4.06623136
NFRキーワード種類 IKn=33	最小レスポンス	1.453279959
	画面レスポンス	4.169152954
	スループット目標値	6.493240967
	レスポンスタイム	3.662394767
リカバリ処理	復旧対応	3.774860382
NFRキーワード種類IKn=53	障害リカバリ手順	1.279856614
	データリカバリ	5.997211406
	リカバリプロセス	3.024930495
	リカバリログ	4.588237967
	回復時間	3.158955849
アクセス管理	不正なアクセス	2.470947366
NFRキーワード種類IKn=134	侵入監視	1.037478679
	不正ログイン	1.79191536
	不正アクセス防止	2.800739317
	アクセス制限	2.36330307
	アカウントロック	3.750156452
負荷バランス	負荷分散	3.7144558
NFRキーワード種類 IKn=41	負荷分散装置	4.509926791
	ロードバランシング	4.494587614
	ネットワーク負荷	3.473900711
	負荷率	3.054600816
セキュリティ管理	セキュリティ管理	2.240994328
NFRキーワード種類 IKn=102	情報セキュリティ対策基準	3.004146602
	セキュリティ管理機能	2.846453825
	セキュリティレベル	3.044960861
	インシデント管理機能	3.40147674
障害対策	障害対策	2.451364807
NFRキーワード種類 IKn=76	システム障害対応	3.464907252
	計画停電対応	3.897737079
	ホットスワップ	3.276547471
	障害検知ソフトウェア	3.033028982
	障害発生頻度	3.520007328

4.4 評価実験

本節では、「NFR キーワード指数」により重み付けされた NFR キーワードによる、RFP の記述内容の明確さを評価する実験について述べる。評価実験では 161 件のサンプル RFP を入力データとしてテキストマイニングにより NFR キー

ワードを抽出した後に、この NFR キーワードを指定してその出現頻度から「NFR キーワード指数」を生成し、RFP ごとに各「小項目」の評価スコアを求めて評価する。次の各項で、評価実験の結果および考察について述べる。

4.4.1 評価実験データと「小項目」

評価実験では、サンプルデータとしてウェブから収集した RFP は text 形式に変換し、データクリーニングを行った後に、各 RFP を一つの文章単位としてすべてのサンプルを含む一つのテキスト文書を作成する。RFP テキストデータの総語数は 4,611,476 語である。また、評価対象とした「小項目」である「応答性」、「負荷バランス」、「リカバリ」、「アクセス管理」、「セキュリティ管理」および「障害対策」に出現した NFR キーワード数はそれぞれ 398 個、467 個、653 個、1,471 個、2,580 個および 823 個であった。

4.4.2 評価実験結果と考察

表 4.3 に各「NFR キーワード指数」による 3 種類の評価手法と「手動評価」との相関係数を示す。表 4.3 によると、相関の強さについてはすべての「小項目」指標について「弱い相関あり」と認められるが、「拡張 TF-IDF」がすべての「小項目」について相関が相対的に強いことがわかる。「小項目」別にみると、「セキュリティ管理」、「障害対策」、「アクセス管理」、「応答性」、「リカバリ処理」および「負荷バランス」の順に相関が強い。上位 3 つの「小項目」については下位 3 つの「小項目」と比較して NFR キーワードの種類数が多いことから、語彙の豊富さが相関の強さに影響していることが考えられる。「拡張 TF-IDF」手法は、対応する「小項目」の各 NFR キーワードの出現頻度が高ければ評価点は高くなるが、同一の NFR キーワードの出現頻度が高くても評価点は高くなり、異なるキーワードの種類が多いほど評価スコアが高くなると言える。しかし、「拡張 TF-IDF」手法の評価スコアが高い場合でも必ずしも「手動評価」が高くないケースがある。(図 4.3 参照)

表4.3 NFRキーワード重みと手動評価の相関係数

手動評価との相関係数	単純TF-IDF	拡張TF-IDF	キーワード比
アクセス管理	0.3267	0.3684	0.3444
セキュリティ管理	0.3344	0.4296	0.4051
リカバリ処理	0.2419	0.2666	0.1869
応答性	0.3352	0.3625	0.325
障害対策	0.3855	0.4113	0.3125
負荷バランス	0.1444	0.2239	0.1451

そこで、161 件の RFP について、個別の「小項目」に関する3つの手法により算出された評価ポイントと3段階の手動評価結果と対応させた結果を図 4.3(a), (b), (c), (d), (e), (f) の箱ひげ図に示す。図 4.3 によると、手動による3段階評価との比較において「応答性」および「負荷バランス」の2つの「小項目」以外は「拡張 TF-IDF」のばらつきが小さいことが認められる。また、「応答性」および「負荷バランス」については「キーワード比」のばらつきが小さいことも認められるが、3つの手法の内「拡張 TF-IDF」手法が最も有効であることを示していると考えられる。ただし、「拡張 TF-IDF」手法の場合であってもすべての「小項目」に外れ値があり、手動評価と NFR キーワードによる評価の判定に大きな差が認められる。人による評価と自動評価に大きな差があるケースとして、「手動評価」が“0”であっても「教師なし自動評価」で高い評価点となっている場合がある。特に、「障害対策」の「拡張 TF-IDF」による評価点は“22.7”であるが「手動評価」は“0”となっている。この場合の NFR キーワードに関連する RFP の記述内容を、表 4.4 に示す。表 4.4 によると「障害対策」という語句についての記述内容には具体的な表現がなく、ただ「提案依頼」を求める内容であることが認められる。また、専門用語としての「ホットスワップ」や「リンクアグリゲーション」については、具体的な記述も認められる。人による RFP の評価の場合には、非機能要件である「小項目」の語句の記述は認めるが、ユーザとしての要求が明確に記述されずに提案依頼を行っていることが、低い評価につながったことが考えられる。さらに、専門用語についての記述については、より具体的な記述が求められるにもかかわらず、簡単な記述であれば評価が低いことが考えられる。人による3段階評価の場合には、「選択バイアス」や、NFR キーワードの出現頻度だ

けに依存しないことが推定される。「手動評価」は、評価者が文章の前後関係を見て判断し、評価していることが考えられる。よって、自動評価と手動評価との評価結果の極端な一致の違いをなくすためには、NFR キーワードの前後関係や文脈を考慮した NFR キーワード指数を生成し、「拡張 TF-IDF」手法を改良することが今後の課題である。

表 4.4 NFR キーワードに関連する記述例 「障害対策」

NFRキーワード	出現回数	記述内容
アクシデントレポート作成機能	1	インシデントレポート・アクシデントレポート作成機能があること。 以下の条件でインシデントレポート・アクシデントレポート作成機能があること。
ホットスワップ	3	各ディスク装置は、RAID構成及びホットスワップ対応等で障害対策が十分に施されていること。 装置内で電源ユニット、ファンモジュールがホットスワップ可能なこと。 装置内で電源ユニット、ファンモジュールがホットスワップ可能なこと。
リンクアグリゲーション	3	上記インタフェースについては、2ポート以上を仮想的に1ポートとする機能（一般的に、リンクアグリゲーション）IEEE 802.3ad リンクアグリゲーション機能（LACP及びスタティック）を有すること。 スタック接続された装置間でリンクアグリゲーション接続が可能なこと。
計画停電	1	計画停電時にタイマーによる自動安全停止及び正常起動を考慮した設計を行うこと。
障害切り分けマニュアル	1	障害切り分けマニュアル
障害対策	1	障害対策 基幹システムは機能単位での分割・分散化を図り、一つの機能の停止が全体システムの停止に波及しないように障害の局所化を図れるものを提案すること。
障害対応	3	想定する障害対応策及び回避策 仕様を満たし、障害原因の追及および障害対応を顧客確認が容易にできるシステム提案であること。 夜間や休日を含む24時間の障害対応について、優れた提案がされていること。
耐障害性	1	さらに耐障害性、性能確保についても十分考慮して総ディスク容量を見積ること。

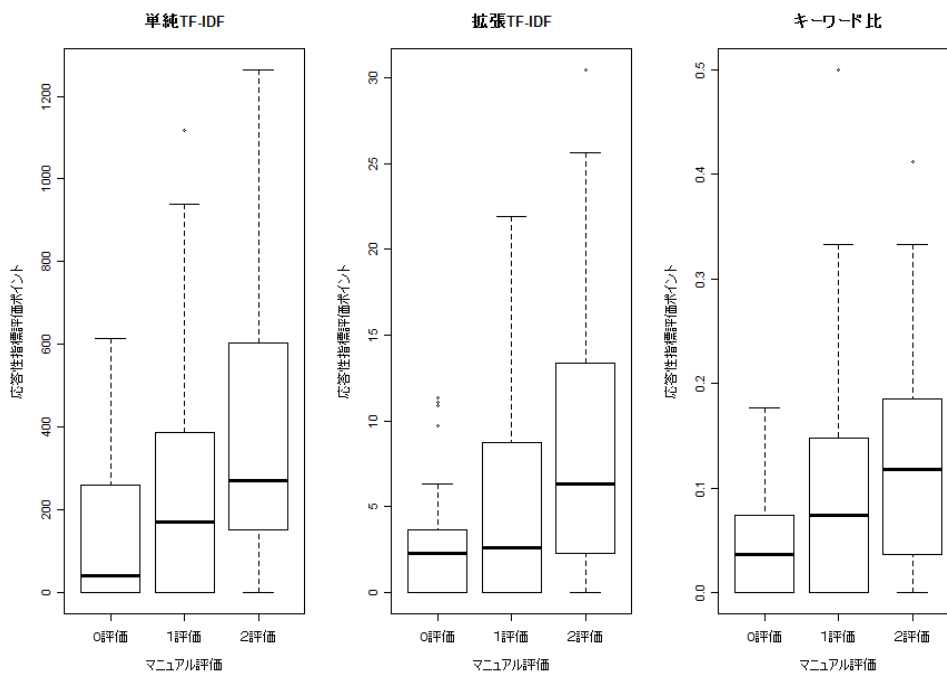


図4.3 (a) 応答性のNFRキーワード評価別箱ひげ図

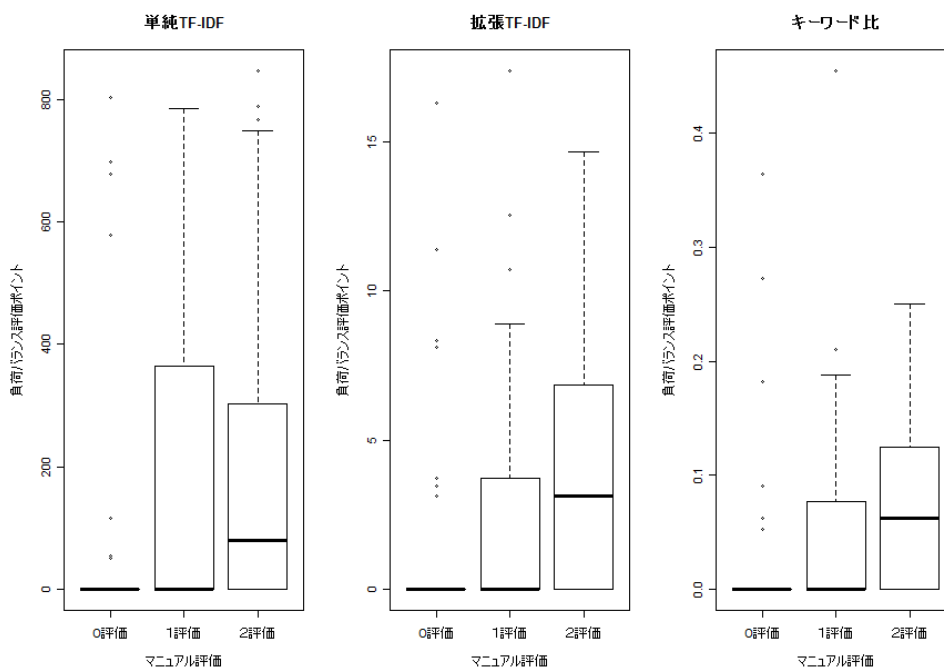


図4.3(b) 負荷バランスのNFRキーワード評価法別箱ひげ図

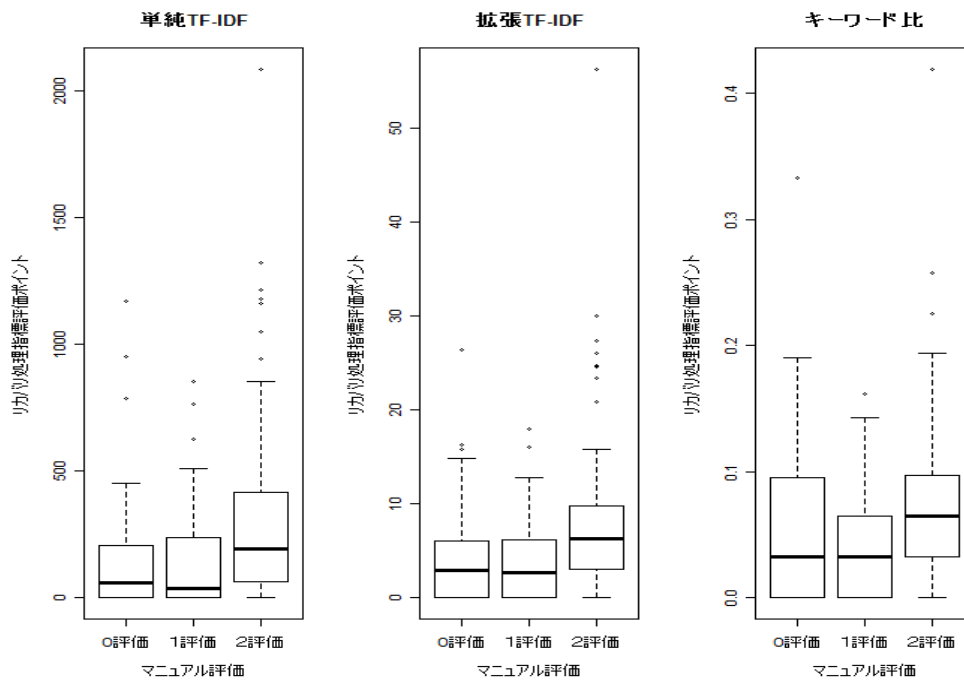


図4.3(c) リカバリ処理のNFRキーワード評価法別箱ひげ図

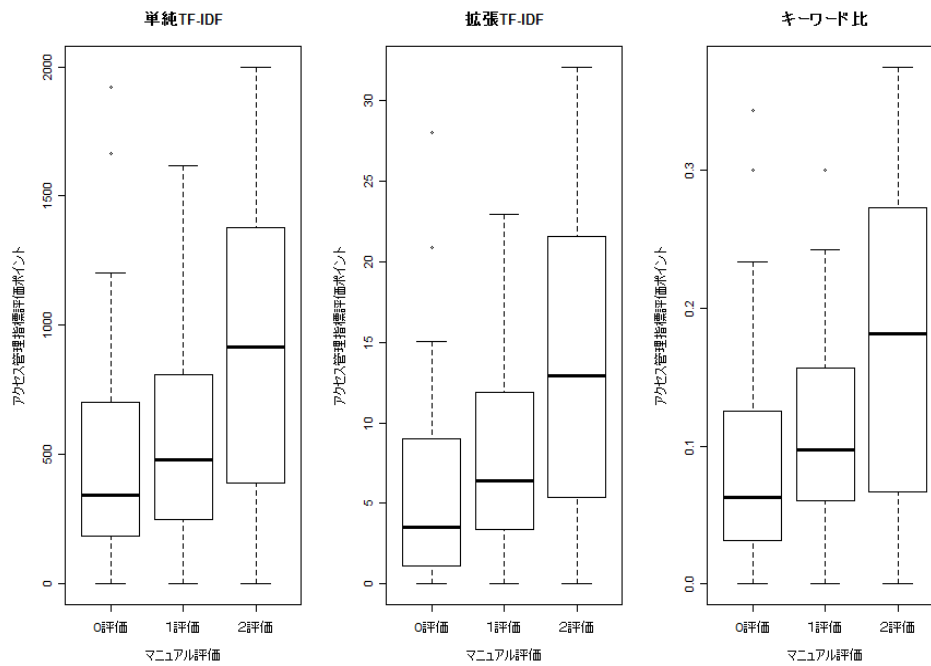


図4.3(d) アクセス管理のNFRキーワード評価法別箱ひげ図

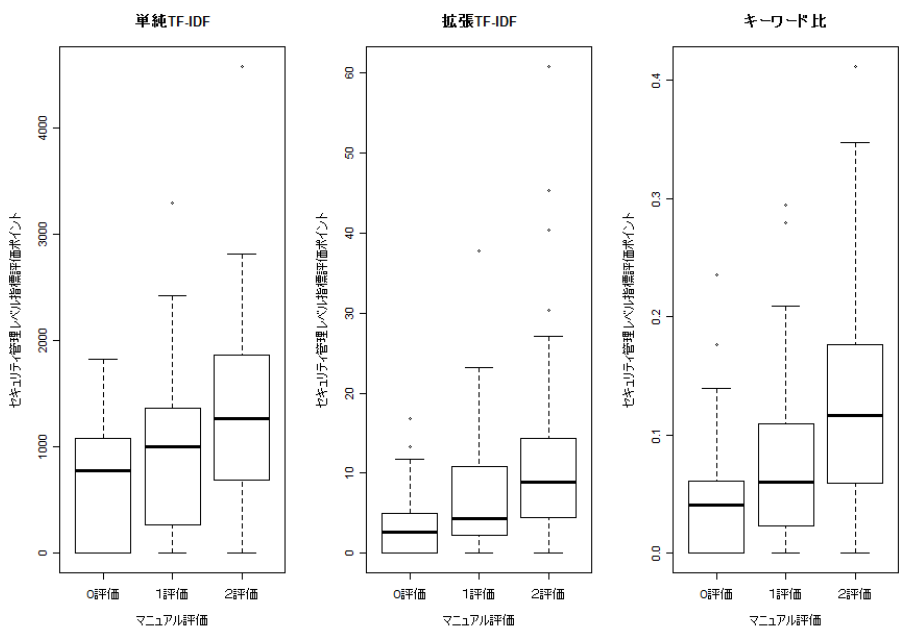


図4.3(e) セキュリティ管理のNFRキーワード評価別箱ひげ図

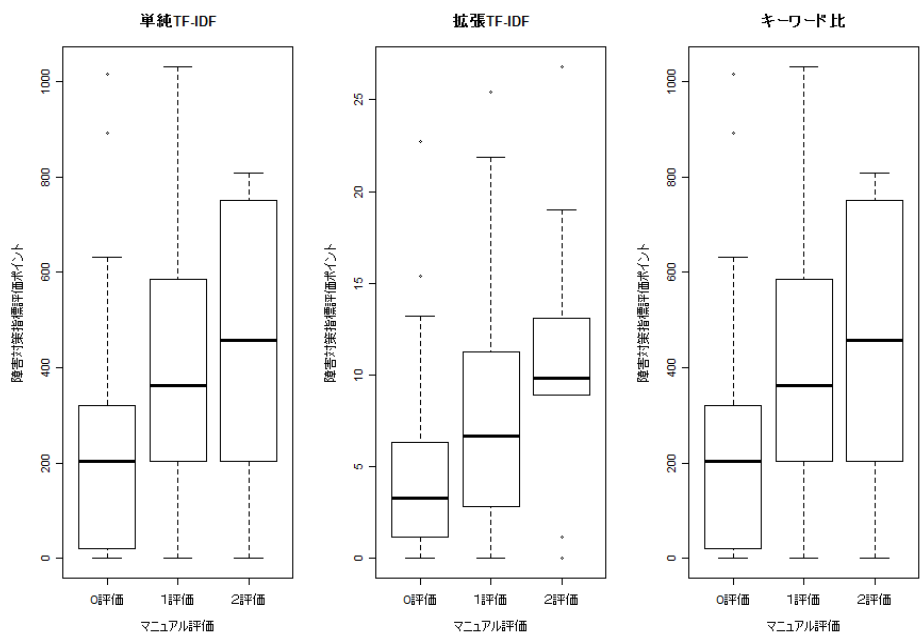


図4.3(f) 障害対策のNFRキーワード評価法別箱ひげ図

図 4.3 NFR キーワード重み付け手法別評価結果の箱ひげ図

次に、6つの「小項目」について、教師なし自動評価手法の拡張 TF-IDF による評価点を「5点未満」、「5点以上 10点未満」および「10点以上」の3区分として、3段階手動評価の「0点」、「1点」および「2点」の各評価に占める比率を算出した結果を図 4.4(a), (b), (c), (d), (e)および(f)に示す。図 4.4(a), (b), (c), (d), (e)および(f)によると「手動評価」が高い場合には、拡張 TF-IDF の評価点も高いことが認められる。このことは、拡張 TF-IDF による教師なし自動評価が多段階評価による「手動評価」の代替となる可能性を示していると言える。しかし、「手動評価」が“2”の場合でも「自動評価」が「5点未満」の評価となる比率が高い「小項目」がある。特に、「負荷バランス」については、自動評価による「5点未満」の評価が65%程度ある。(図 4.4(b)参照) これは、「負荷バランス」の NFR キーワード数が比較的少ない41個であること、および出現頻度が少ないために教師なし自動評価の評価点が低くなったと考えられる。

一方、「応答性」については、NFR キーワード数が33個と「負荷バランス」より少ないが、出現頻度が高いため、自動評価の評価点は比較的高く、「手動評価」が“2”の場合の「10点以上」の比率が35%程度となっている。これは、人による「手動評価」では「小項目」毎に3段階評価を行い、NFR キーワードの出現頻度にかかわらず記述が明確であれば高評価していることによることが考えられる。また、図 4.5 に56件の RFP について、例として、「アクセス管理」の「教師なし自動評価」と「手動評価」の NFR キーワードの出現回数の結果を示す。図 4.5 によると、「教師なし自動評価」と「手動評価」の不一致については以下のようないふケースがあることが認められる。

- ①NFR キーワードの出現頻度が20を超え評価が高い場合であっても、「手動評価」では“0”評価のケース
- ②NFR キーワードの出現頻度は10を超え、評価点が中程度の場合であっても、「手動評価」では“0”のケース
- ③NFR キーワードの出現頻度が10以下で低く、評価点も10以下で低い場合であっても、「手動評価」では“2”のケース

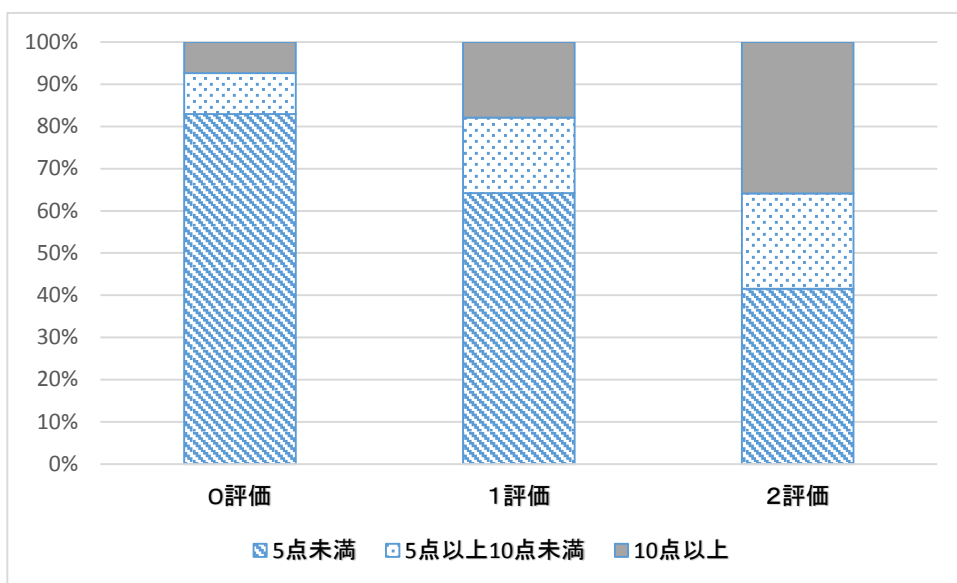


図4.4(a) 手動評価点毎の拡張TF-IDF評価点の比率:応答性

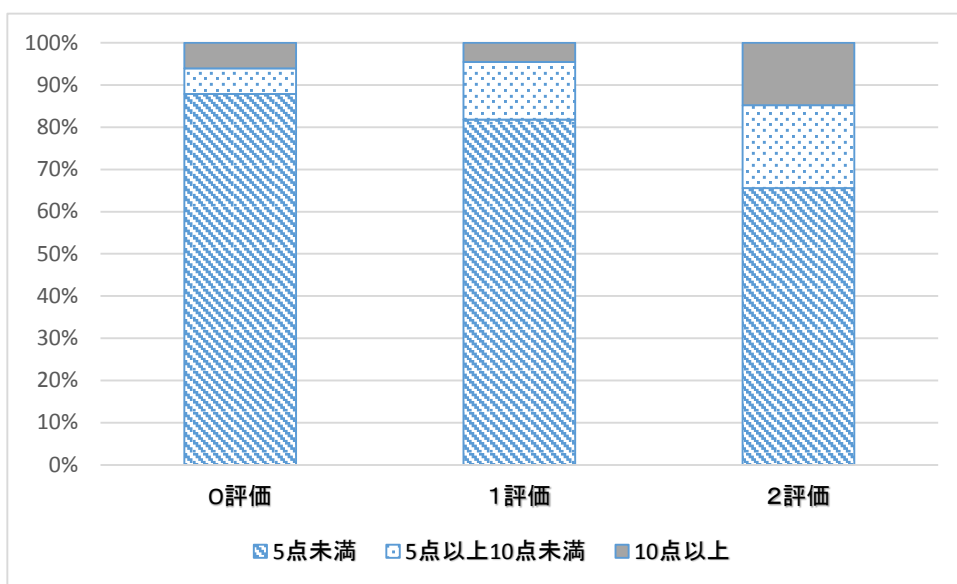


図4.4(b) 手動評価点毎の拡張TF-IDF評価点の比率:負荷バランス

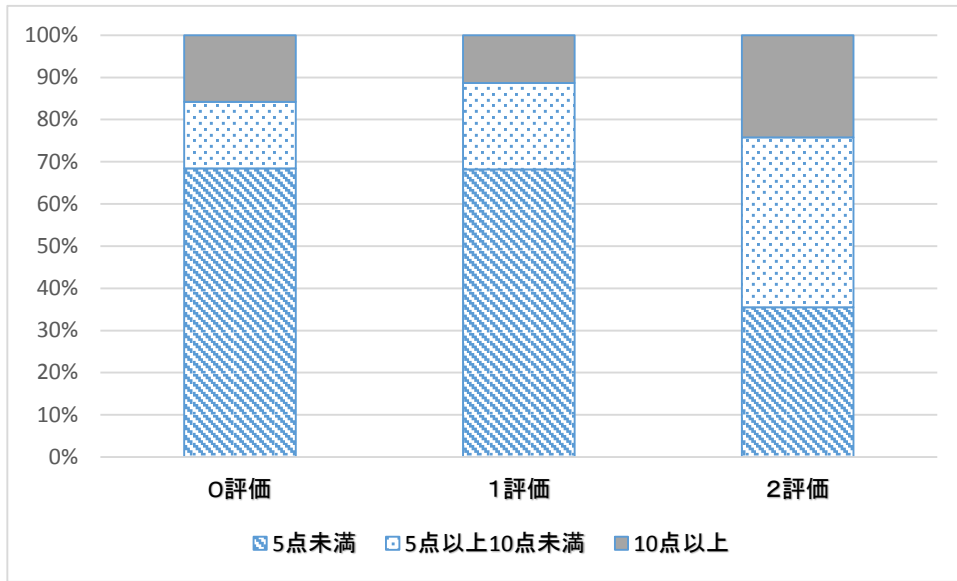


図4.4(c) 手動評価点毎の拡張TF-IDF評価点の比率:リカバリ処理

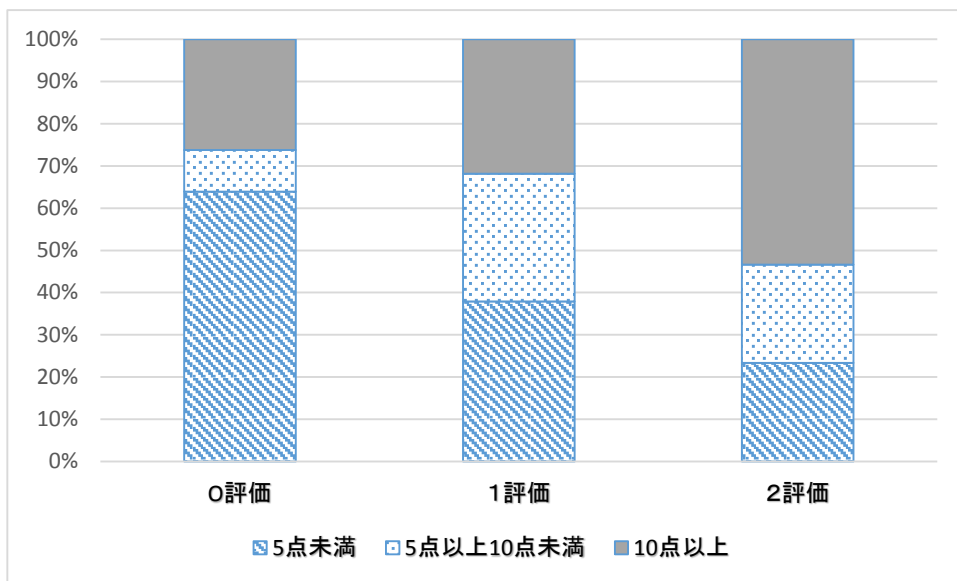


図4.4(d) 手動評価点毎の拡張TF-IDF評価点の比率:アクセス管理

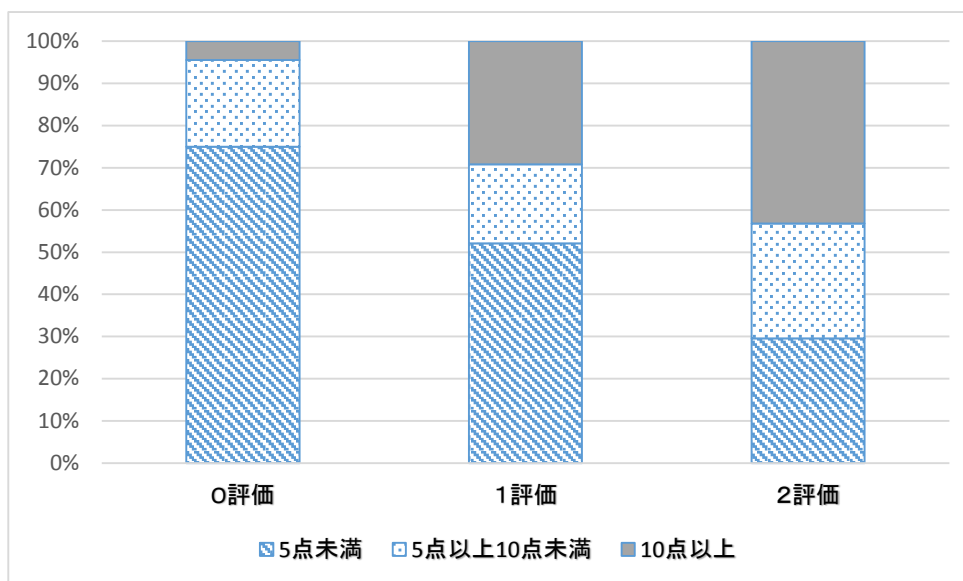


図4.4(e) 手動評価点毎の拡張TF-IDF評価点の比率:セキュリティ管理

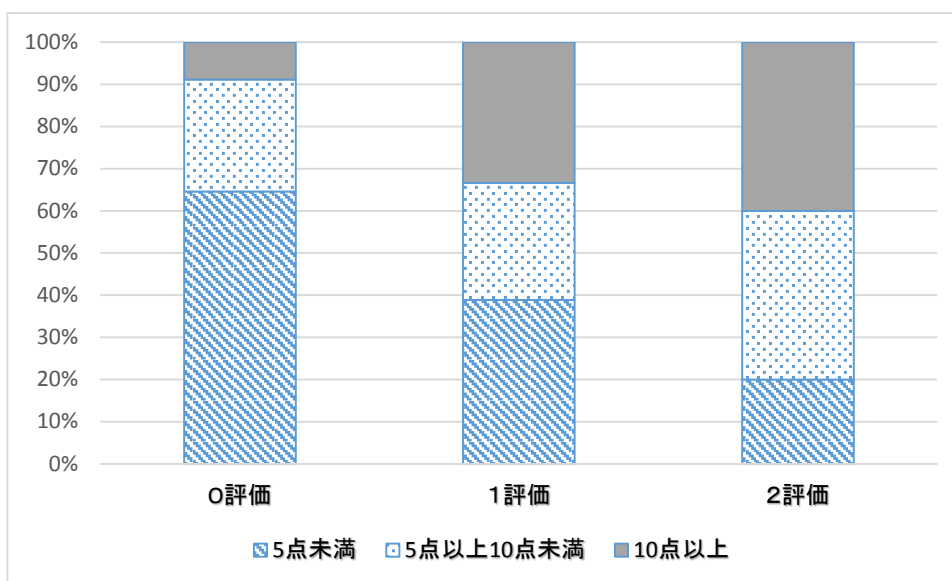


図4.4(f) 手動評価点毎の拡張TF-IDF評価点の比率:障害対策

図 4.4 手動評価点毎の拡張 TF-IDF 評価点の比率

さらに、一部の「小項目」（アクセス管理、リカバリ処理および障害対策）についてはマニュアル評価が“0”であっても拡張 TF-IDF の評価点が「20 点以上」の高評価の RFP もある。しかし、図 4.6 に示すように手動 3 段階評価別の拡張 TF-IDF の評価スコア平均値は段階評価との

結果を明確に示していることが認められる。これらの結果は、非機能要件の記述の明確さを評価する手法として、教師なし自動評価手法を手動評価の代替として使用することは難しいが、「手動評価」との補完的な使用として有効であると考えられる。今後は、NFRキーワードの前後関係を含めた文脈を「教師なし自動評価手法」に反映させる等により、「手動評価」との一致率を高めることが課題である。

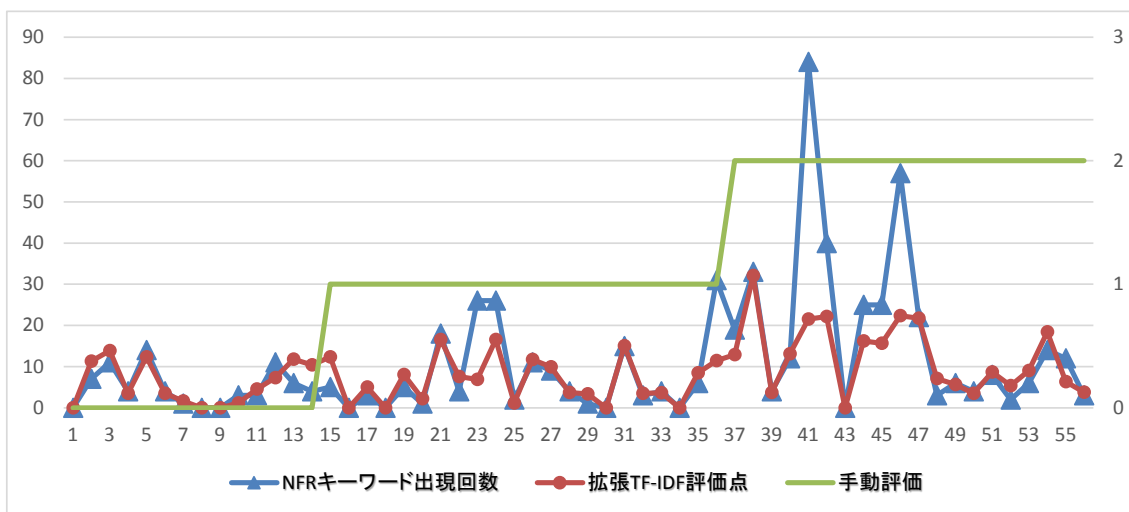


図4.5 手動評価別のNFRキーワード出現回数と拡張TF-IDF評価点

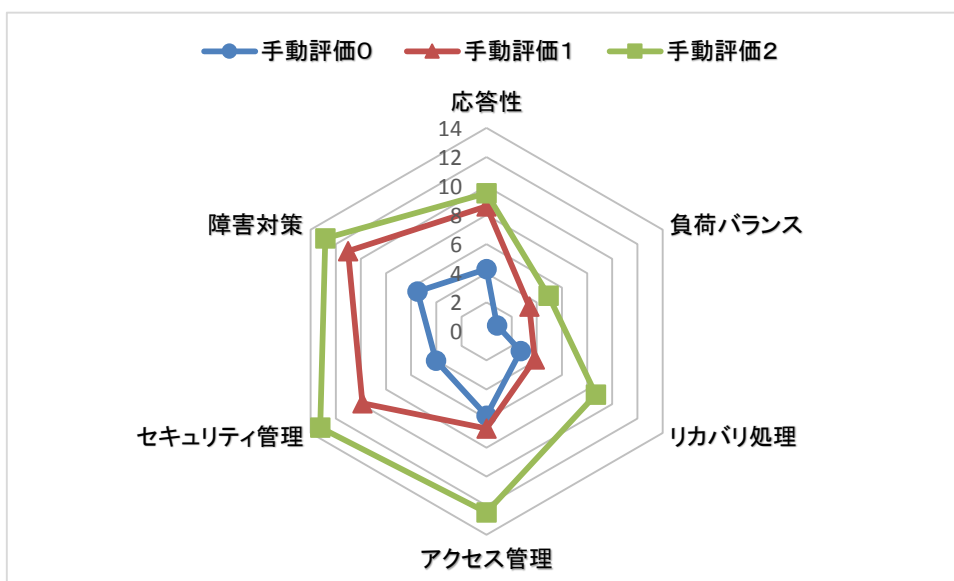


図4.6 手動3段階評価別の拡張TF-IDF評価スコア

4.5 クラスタリングによる評価スコアの段階評価

段階評価により, RFP の非機能要件の記述内容が個々の「大項目」, 「中項目」および「小項目」についてどの程度明確であるかを相対的に把握することが可能であり, 超上流工程においてユーザにとって有用な指標となる. しかし, RFP の評価スコアを算出し, 前節で提案した 3 つの手法により, 段階評価による分類を自動評価することには課題がある. そこで, 追加実験として, 教師なし機械学習の一手法であるクラスタリングにより, 評価スコアに応じて 3 段階評価に分類する実験を行った. クラスタリング手法を用いることにより, RFP から抽出した NFR キーワードと関連付けた「小項目」について, 定量的に評価した結果が類似性により分類されることが推定される.

クラスタリングについては, 文献[32]によると「互いに関連する文書同士は, 関連しないものよりも類似している」というクラスタ仮説 (cluster hypothesis) に基づいている. 前節では「拡張 TF-IDF」が手動評価との比較において, 3 つの手法の中では相関が高く, 評価も類似していると考えられる. また, 「キーワード比」は NFR キーワードの豊富さを示す指標であることから, 3 グループへのクラスタリング実験では, 「拡張 TF-IDF」と「キーワード比」の 2 次元データによる k-means 法を用いた. 「小項目」の 3 段階評価に対応して評価するために, 3 グループにクラスタリングされた結果と, 評価者が各 RFP について「手動評価」した結果との一致率について検証を行う. 次項では提案手法の詳細について述べる.

4.5.1 提案手法

「小項目」ごとに「拡張 TF-IDF」により算出した NFR 評価スコアと「キーワード比」のペアを列とし, 各 RFP を行として, k-means 法を適用する.

K-means アルゴリズムは「非階層的クラスタ分析であり, クラスタ数を指定してその個数のクラスタを構成する手法」である. 文献[40]では, 「k-means アルゴリズムは, ベクトル集合 $\{x_1, x_2, \dots, x_n\}$ を一連のクラスタ $\{C_1, C_2, \dots, C_k\}$ に分割する. こ

のアルゴリズムは、はじめにシードと呼ぶ k 個の初期クラスタ要素を必要とする。これらは外部から与えられるか、またはベクトルの中からランダムに選択される。」とされている。また、「**k-means** アルゴリズムはシンプルかつ効率的なため、頻繁に用いられる。各々の反復処理においては $O(kn)$ 回の類似度比較を行えば良く、通常、収束するまでの反復処理の回数は非常に少ない。」と言われている。しかし、「**k-means** アルゴリズムにおける大きな問題点は、得られる結果が最初のシードの選択に依存すること」であり、この問題を解決するために、クラスタリングを複数回実行し、その都度異なったシードを選ぶ方法をとっている。本項では、クラスタ数 3 を指定してクラスタリングを複数回実行することにより、RFP 群の評価結果を 3 つのクラスタとして構成し、「小項目」ごとに 3 段階評価との対応付けを行っている。

クラスタリングの評価[40]には 4 種類あるが、最も一般的な尺度は、ある正解のクラスタのデータを、どの程度含むかという純度(purity)である。人によって分類された正解となるクラスタを $\{A_1, A_2, \dots, A_n\}$ とし、 $\{C_1, C_2, \dots, C_m\}$ をクラスタリング手法によって得られたクラスタとすると、純度は以下の式で定義される。純度は 0 から 1 の値をとり、値の結果が高いほどクラスタリングの結果が良いとされている。

$$\text{Purity}_i = \frac{1}{|C_i|} \text{Max}_h |C_i \cap A_h|$$

また、もう一つの一般的な尺度として以下に示すエントロピー尺度がある。エントロピーは 0 から 1 の値をとり、値が低いほどクラスタリング結果が良いとされている。

$$\sum_{i=1}^K \frac{|C_i|}{N} E_i = - \sum_{i=1}^k \frac{\sum_{j=1}^K x_{ij}}{N} E_i$$

ここで、 N はクラスタリング対象のデータ数、また E_i は以下で定義される。

$$E_i = - \sum_{h=1}^K P(A_h | C_i) \log P(A_h | C_i)$$

また、確率 $P(A_h | C_i)$ は $\frac{|A_h \cap C_i|}{|C_i|} = \frac{x_{ih}}{\sum_{j=1}^K x_{ij}}$ によって推定される。

4.5.2 クラスタリング評価実験

NFR キーワードを抽出した後、クラスタリングによる評価実験を行う手順を以下に述べる。

手順 1: 「拡張 TF-IDF」による「NFR キーワード指数」で算出した NFR 評価スコアと RFP 群のキーワード比とのペア数値をクラスタリング対象データとして、クラスタ数 3 を指定して非階層クラスタ分析手法である k-means を実行する。k-means 法によるクラスタリングは最初にランダムに選択されたシードに依存することから、試行回数 50 回を指定して実行する。

手順 2: 3 グループにクラスタにクラスタリングされた結果から、マニュアルにより NFR 評価スコアの大小関係をもとにスコアの高いクラスタの評価を“2”，スコアが中間のクラスタの評価を“1”およびスコアが低いクラスタの評価を“0”に対応づける。

手順 3: クラスタ分析による評価結果と手動評価結果の一致率について分析を行う。クラスタリングの評価結果とマニュアル評価のクロス表を作成し、クラスタリング評価尺度を求める。

4.5.3 評価実験結果と考察

「小項目」ごとのクラスタリング結果と手動評価との比較を図 4.6 に示す。図 4.6 によると、すべての RFP について、「小項目」のクラスタリング結果と手動評価における一致率(3 段階評価の結果が“2”，“”および“0”で一致している比率)は「小項目」の平均値で 60.7%であった。また、不一致率は「小項目」の平均値で 38.5%であった。一致率 60.7%の内訳は、評価スコア“2”で一致が 10.2%，評価スコア“1”で一致が 24.7%および評価スコア“0”で一致が 65.0%となっている。一方、不一致率 38.5%の内訳をみると評価スコアの“2”を“1”あ

るいは“1”を“2”とした評価が 32.7%であり、評価スコアの“1”を“0”あるいは“0”を“1”とした評価が 67.3%であった。評価が一致した場合および不一致であった場合のいずれも、「小項目」の評価スコアが低い“0”が多かったことが、これらの要因と推定される。また、RFP の「小項目」にもとづく非機能要件の記述内容の明確さの評価において、「該当する指標に関する記述がない、あるいは記述はあるがその内容が極めてあいまいな表現である」RFP が多いことも示している。しかし、評価スコアの“2”を“0”あるいは“0”を“2”とした評価は10個であり、161件のRFPを6種類の「小項目」で評価した966ケースの0.9%であったことは、クラスタリングによる3段階評価の精度向上の可能性を示していると言える。

一方、クラスタリング評価の「小項目」全体について、評価尺度であるエントロピー及び純度で見ると、それぞれ 0.737 と 0.606 であった。クラスタリングの結果はエントロピーが高く、純度も低いため良いとは言えないが、委託ソフトウェア開発の超上流工程において、評価スコアを大まかに3段階に分類し、どの程度の評価にあるかを直観的に把握することに効果があると考えられる。そこで、不一致となった要因を個別に考察することにより、今後の課題を明らかにする。

「小項目」ごとの一致率および不一致率を図 4.7 に示す。図 4.7 によると、「負荷バランス」の一致率が最も高く、「応答性」の一致率が最も低い結果が認められる。「負荷バランス」について一致率が高いことの一つの要因として、評価スコアの“0-0”及び“1-1”の一致数が多いことが考えられる。「応答性」について一致率が低い要因の一つとして、レスポンスの記述と具体的数値が記述されている RFP について、手動評価は“2”であるがクラスタリングでは“0”評価となっていることが一つの要因と考えられる。この例では、「応答性」についての定量的要求が明確であれば手動評価としての評価スコアが高いが、NFR キーワードの出現頻度が少ないため評価スコアが低くなり、k-means 法によるクラスタリングでは“0”評価となっていることが推定される。そこで、手動評価とクラスタリング結果の一致率に差異が生じる要因および解決策について以下に考察する。

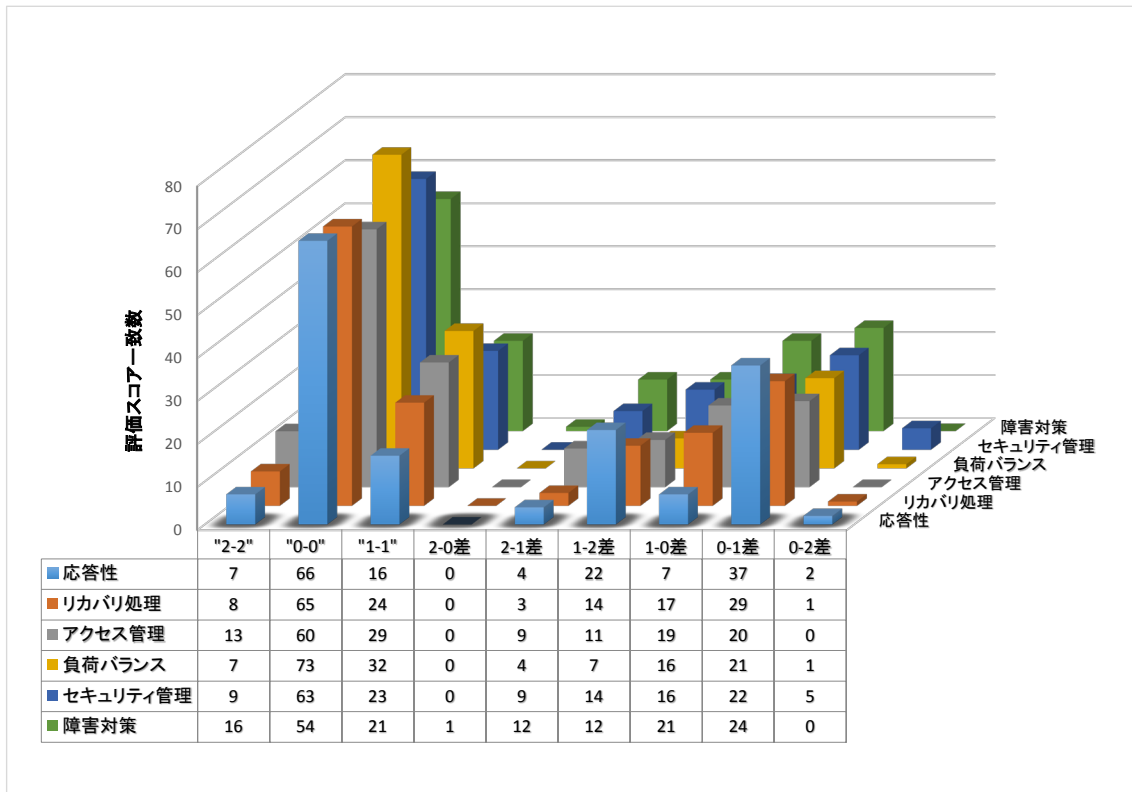


図 4.7 小項目ごとの NFR 評価スコア一致数

A) 手動評価が高く、クラスタリング評価が低いケース

複合語としての専門用語が記述されていないが、平易な語句で該当する指標の要求を記述しているケースが考えられる。例として、「リカバリ処理」にて、「システムに障害が発生した場合は 2 時間以内に対応し、障害箇所の究明を行ったうえで、応急的措置」と記述されているものがある。ここでは「システム障害」としての複合語ではなく「システムに障害」と記述されており NFR キーワードとしていないため評価されない。また、手動評価では「2 時間以内に対応」を定量的記述として評価するが NFR キーワードではないので評価ポイントがなく、評価スコアは低いためクラスタリングでの評価は低くなる。さらに、RFP に記述されている単語表現に問題があり、テキストマイニングによる形態素解析で抽出できない場合があった。例として、「レスポンス」を「リスポンス」と記述、「画面応答」を「画面レスポンス」と記述しているなどである。これらの問題点に対して、文脈

解析および意味解析の手法を取り入れること、あるいは評価する NFR キーワード表現の多様性を許容することが考えられる。

B) 手動評価が低く、クラスタリング評価が高いケース

複合語としての多様な専門用語が記述されているが、該当指標に対する体系的な記述がないケースがある。これは、RFP の記述として NFR と FR を明確に区分せずに FR の中に NFR を部分的に記述している場合である。さらに、同一フレーズを繰り返して記述している場合に、該当プロジェクトでの複数の特定 NFR キーワードの出現頻度が極めて大きくなり、「NFR 評価指数」による評価スコアが高くなる場合である。これは、NFR キーワードが文節の表題とされている場合、あるいは RFP の記述において文章の繰り返しを多用している場合である。これらの問題点に対して、サンプルデータである RFP 原文を加工して NFR 記述部分をまとめること、あるいは繰り返されている NFR キーワードを減少させる手法が必要となる。また、“±1”の評価差が生じる要因として以下のケースが考えられる。

C) クラスタリングと手動評価が“0-1”となるケース

NFR 評価スコアが比較的低い RFP 群のクラスタリングは“0”となったが、手動評価では、特定の NFR キーワードの重要性を判断し評価を“1”としているケースである。これに対しては、NFR キーワードの多様性を高め NFR キーワード数を増加することにより評価スコアを上げる方法が考えられる。しかし、「出現頻度で上位 10%の単語しか使わなくても、分類器の性能は下がらないという実験の結果がある」[32]とされており、少ない NFR キーワードで分類を可能とすることが課題となる。

D) クラスタリングと手動評価が“1-2”となるケース

評価スコアが極めて高いプロジェクトが少数あり、これに影響されて少数のプロジェクトのみが“2”に分類され、手動評価では“2”と判断されるケ

ースで、クラスタリングでは中間評価の“1”となるケースである。これに対しては、「小項目」の評価スコアが極めて高い RFP を外れ値として処理することにより NFR 評価に差異が少ないサンプルデータを対象とすることにより問題点を避けることができる。ただし、「小項目」の評価スコアが高い RFP を対象外とすることは、非機能要件の明確さを評価するためには問題であり、外れ値の取り扱いは今後の課題である。

E) クラスタリングと手動評価が“1-0”となるケース

NFR キーワードではないにもかかわらず、NFR キーワードとして評価ポイントとなった場合である。たとえば、「負荷検査」は医療機関における検査項目として使用されるケースがあり、「病院情報システム」に記述される、あるいは“ピーク時”は業務そのもののピーク時間帯を意味する語句として使用されるなどである。これは、NFR キーワードが、実際は非機能要件とは異なる意味のために、手動評価では評価されないことが考えられる。これに対しては、自然言語処理の手法による語義のあいまい性解消やキーワードのコロケーション統計(コロケーションは一般に語の習慣的な共起関係と説明されている)を評価手法に組み込むことなどが対策として考えられる。

F) クラスタリングと手動評価が“2-1”となるケース

重みが高い NFR キーワードが記述されているが、その記述内容が簡単であり、手動評価ではあまり評価されないケースである。たとえば、「障害管理」指標としての「障害管理」、「障害復旧」あるいは「障害対策」は比較的「NFR キーワード指数」の評価が高いが、「障害が発生した場合は、障害の復旧作業を行うこと」などと平易に記述されているなどである。これは、あまり専門的ではない NFR キーワードであるが RFP サンプル群での出現頻度が高いため「NFR キーワード指数」が比較的高いことによることが考えられる。これに対しては、専門性が高い NFR キーワードについては「NFR キーワード指数」の重みを大きくする手法、あるいは一般的な

NFR キーワードをフィルタリングすることなどが対策として考えられる。

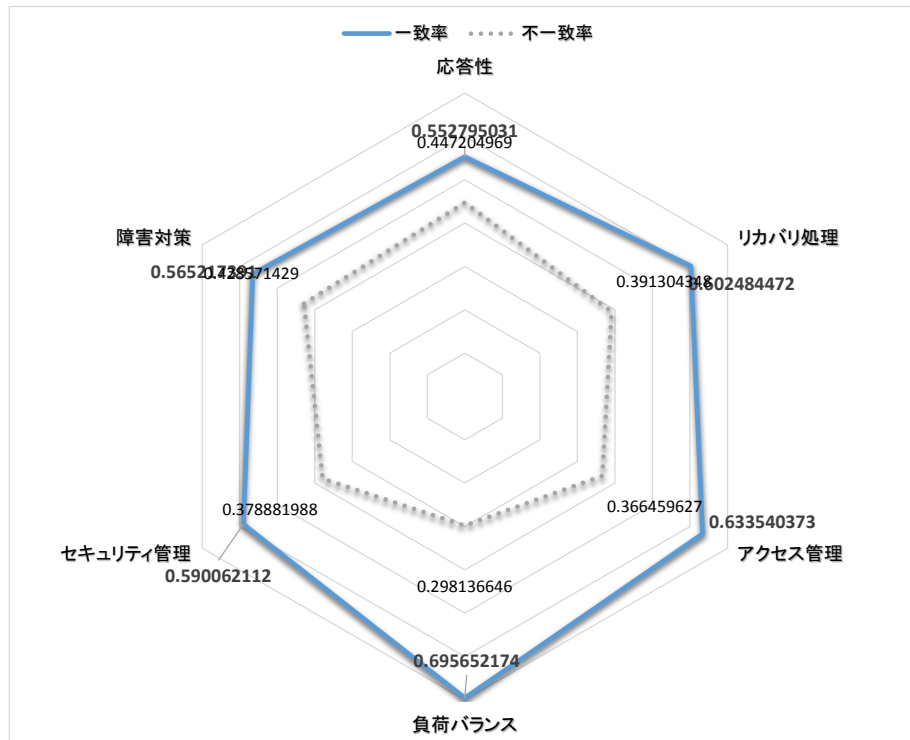


図4.8 「小項目」別の一致率および不一致率

4.6 まとめ

本章では、機械学習によらない非機能要件の評価手法として提案した 3 つの手法を比較した。実験結果では、「手動評価」との各「小項目」についての比較で、「拡張 TF-IDF」手法の相関係数が 0.2239 から 0.4296 となり、「弱い相関」であるが、最も有効であることがわかった。しかし、「拡張 TF-IDF」手法の評価スコアが高い場合でも必ずしも「手動評価」が高くないケースもあり、教師なし自動評価手法を手動評価の代替として使用することは難しく、「手動評価」との補完的に使用することが考えられることがわかった。また、「拡張 TF-IDF」と「手動評価」との比較において、各「小項目」に外れ値となった RFP があり、NFR キーワードによる評価に文章の前後関係や文脈も含めて評価することが必要であり、今後の課題であることもわかった。

一方、追加実験で行った「手動評価」とクラスタリングについては、その一致率は 60.5%であった。さらに、不一致率は 0.9%であった。提案手法の精度評価において、クラスタリングの評価尺度であるエントロピー及び純度はそれぞれ 0.737 と 0.606 であった。エントロピーを高めている要因としては、評価差の違いが±1 であるケースが 38.2%と比較的多いことが考えられる。この点については、RFP サンプルデータとして実験を行った 161 件の RFP データ群の「小項目」における評価が低いデータが多いことも要因のひとつであり、今後は、この±1 差の評価を減少させる手法を組み込むことが課題である。また、手動評価を正解集合としてクラスタリングの結果と一致率の評価比較を行ったが、手動評価の客観性を高めることも必要であり、多数の評価者による評価を行った結果を採用した実験を行い、検証する必要がある。

第 5 章 社会的影響別の非機能要件評価

5.1 概要

非機能要件は、情報化対象となっているシステムのカテゴリ、あるいは規模により重視される非機能要件とそれほど重視されない非機能要件があると推定される。しかし、委託開発を行う情報化システムの規模や対象領域は様々であるが、社会的影響が同じカテゴリの中での比較、あるいは、異なるカテゴリとの比較を行い、どの「非機能要件」の記述が明確でないかを定量的に把握し、確認できることは有効である。そこで、本章では、社会的影響の違いによる 3 つのカテゴリに分類した RFP 群について、前章で述べた「拡張 TF-IDF」手法による評価を行い、「NFR キーワード評価シート」による非機能要件の記述内容の明確さを評価することの有効性について述べる。

「社会的影響によるシステム分類の判定基準」については、IPA-SEC の「非機能要求グレード」[2]で以下のように定義されている。

社会的影響によるシステム分類の判定基準

A) 社会的影響が極めて大きいシステム： 常時サービスの提供が前提となるため、基本的に停止は不可。

このカテゴリの NFR 評価指標による評価が最も高いことが推定される。

B) 社会的影響が限定されるシステム： 24 時間無停止でのサービス提供は必要ないが、極力稼働は継続させる。

このカテゴリの NFR 評価指標による評価は中間であることが推定される。

C) 社会的影響がほとんど無いシステム： 業務時間内のみのサービス提供であり、夜間の稼働は無い。

5.2 社会的影響による RFP の分類と手動評価

本章で対象とした 161 件の RFP サンプルを「社会的影響によるシステム分

類の判定基準」により「社会的影響が極めて大きいシステム」、「社会的影響が限定的なシステム」および「社会的影響が少ないシステム」3つのカテゴリに分類した。RFP サンプル群は、社会的影響別にそれぞれ 56 件、56 件および 49 件に分類された。

これらについて人手による評価を行った「小項目」ごとの評価スコアを図 5.1 に示す。図 5.1 によると「社会的影響が極めて大きいシステム」の評価は「社会的影響が低いシステム」との比較では、すべての「小項目」において高い評価となっており、「社会的影響が限定的システム」との比較においては、「ウイルス対策」、「RAID 構成」および「障害対策」が低い評価であるが、その他の「小項目」においては、すべて高い評価となっていることが認められる。このことは「NFR キーワード評価シート」の「小項目」の手動評価において、社会的影響カテゴリの特徴を評価できることを示していると考えられる。社会的影響カテゴリに分類された個別の RFP についての 3 段階での手動評価結果の一部を表 5.1 に示す。表 5.1 によると、3 段階評価において特に「セキュリティ管理」と「障害対策」の“0”評価がその他の「小項目」と比較して多く、非機能要件の記述においてさらに明確にする必要があることが認められる。また「リカバリ処理」、「応答性」および「負荷バランス」については“2”評価の RFP が相対的に多いことから、「社会的影響が大きいシステム」ではこれらの「小特性」についての非機能要件が重視されていると言える。一方、「社会的影響が少ないシステム」では“0”評価の RFP が相対的に多く、非機能要件について明確に記述されていないことが認められる。ただし、これらについては評価対象とした 161 件の RFP を分類した評価結果であり、複数評価者によるより多くの RFP について分析し、評価することが求められる。次節では、4 章で述べた「拡張 TF-IDF」を用いた評価について述べる。

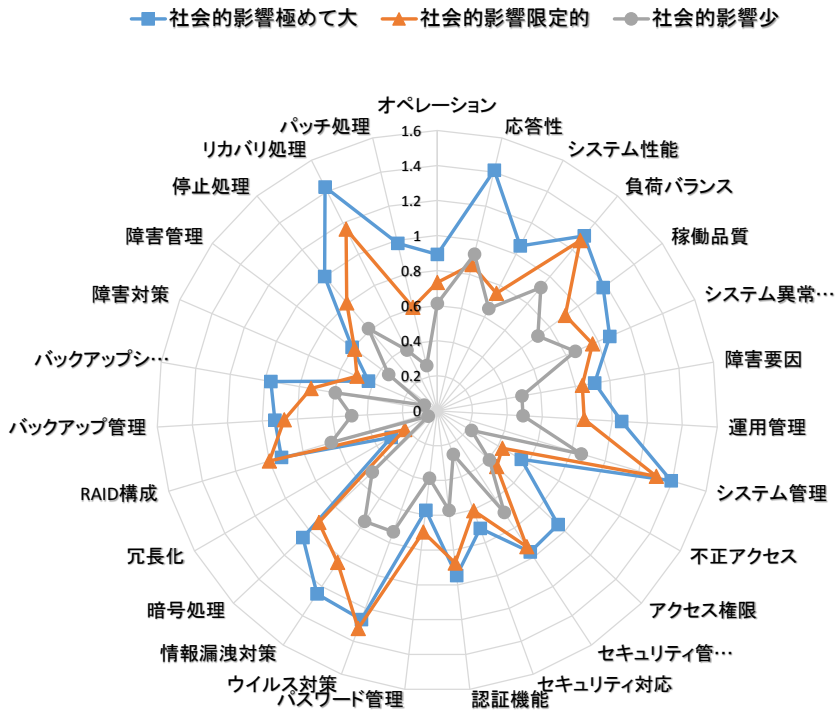


図 5.1 社会的影響別の「小項目」の手動評価スコア

表 5.1 社会的影響カテゴリ別の手動評価(一部の NFR 評価指標)

アクセス管理	影響大	影響限定	影響少	セキュリティ管理	影響大	影響限定	影響少
0評価	14	25	27	0評価	21	21	25
1評価	22	25	19	1評価	16	18	10
2評価	20	6	3	2評価	19	17	10

リカバリ処理	影響大	影響限定	影響少	障害対策	影響大	影響限定	影響少
0評価	14	11	32	0評価	37	33	45
1評価	4	25	15	1評価	14	18	4
2評価	38	20	2	2評価	5	5	0

応答性	影響大	影響限定	影響少	負荷バランス	影響大	影響限定	影響少
0評価	10	19	13	0評価	8	9	16
1評価	14	26	27	1評価	23	23	21
2評価	31	11	9	2評価	25	24	12

5.3 社会的影響カテゴリ別の評価スコア

4章で提案した「拡張 TF-IDF」法を用いて、6つの「小項目」について社会的影響カテゴリ別に分類された個別の RFP の評価を行った結果を図 5.2(一部抜粋)に示す。図 5.2 によると、どの「小項目」の評価スコアが高く、またどの「小項目」の評価スコアが低いかについては、社会的影響カテゴリ別に分類された個別の RFP で大きく異なることが認められる。「社会的影響が極めて高い」カテゴリの RFP についても、RFP 個別にみると評価スコアが低い RFP(たとえば「アクセス管理」についての 3, 4, 42 番目の RFP など)があり、その「小項目」についての非機能要件が明確に記述されていないことが認められる。また、「社会的影響が限定的」なカテゴリにおいても、非常に高い評価スコアの RFP(たとえば RFP「アクセス管理」および「セキュリティ管理」の 49 や 52 番目の RFP)がある。一方、「社会的影響が少ない」システムであっても「アクセス管理」や「応答性」および「負荷分散」について高い評価スコアとなっている RFP(47 番目)があることが認められる。これらのことは、社会的影響によらず一部の非機能要件である「小項目」については個別の RFP に差があることを示している。

次に、図 5.3 に 6 つの「小項目」について社会的影響カテゴリ別に分類された RFP の評価スコアの箱ひげ図を示す。図 5.3 によると個別の「小項目」についての評価スコアをカテゴリ別にみると、「応答性」と「負荷分散」については「社会的影響が極めて大きいシステム」の中央値が高いことが認められるが、その他の「小項目」については「社会的影響が限定的なシステム」との差は少ない。また、「社会的影響が限定的なシステム」については、「セキュリティ管理」と「リカバリ処理」に外れ値として極めて高い評価スコア(50 ポイント超え)の RFP が認められる。この点については、社会的影響によるカテゴリの分類において、一部の RFP(たとえば 49 番目「厚生労働省 ネットワークシステムの更改」や 52 番目「職業紹介サブシステム統合運用監視業務」)の社会的影響が限定的であるか、あるいは極めて大きいかの判断が困難なため、人の判断によるカテゴリ分類の要因が考えられる。また、「社会的影響が少ないシステム」についてはすべての「小項目」において低い中央値(5 ポイント以下)となっており、社会的影響

による3つのカテゴリの中では、非機能要件の記述内容が明確でないことがわかる。ただし、すべてのカテゴリにおいて評価スコアの中央値が10ポイント以下であった。対象となった161件のRFPに偏りがあることも考えられ、RFPのサンプル数に要因があるかどうかについての検証は、さらに多くの多様なRFPによる実験が今後の課題である。次節では、「手動評価」との比較について述べる。

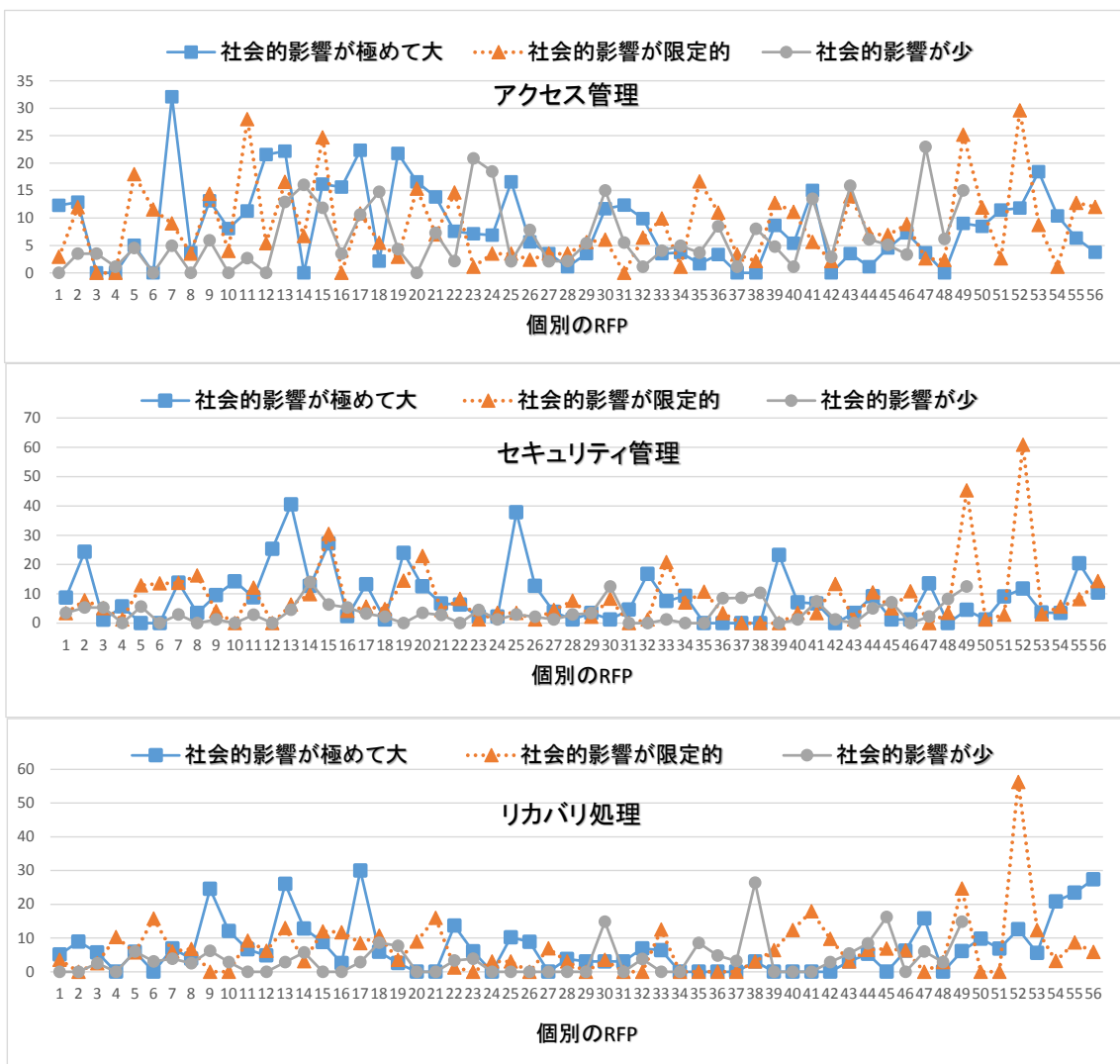


図 5.2(a) 社会的影響カテゴリ別の個別 RFP の小項目評価スコア

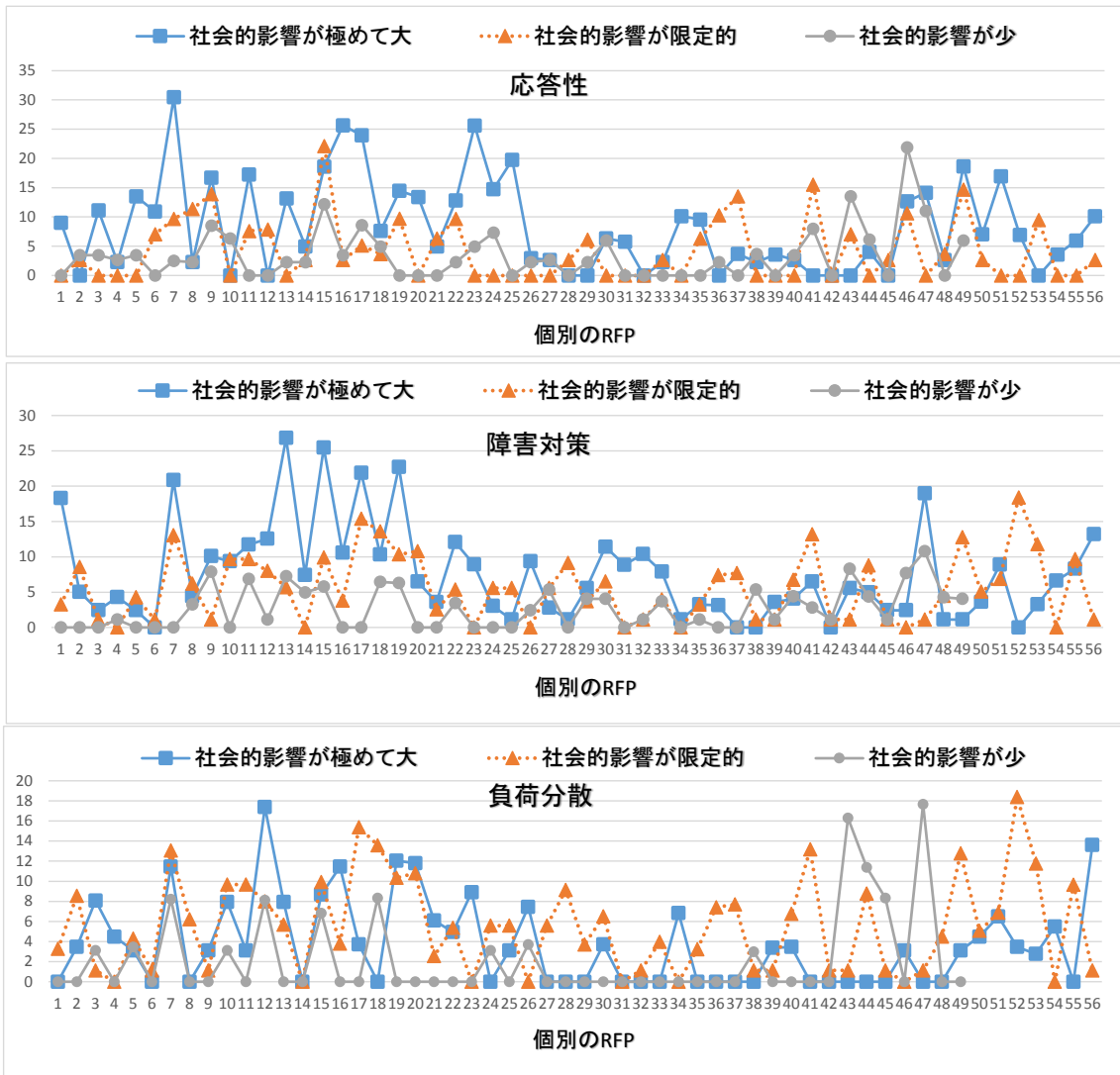


図 5.2(b) 社会的影響カテゴリ別の個別 RFP の小項目評価スコア

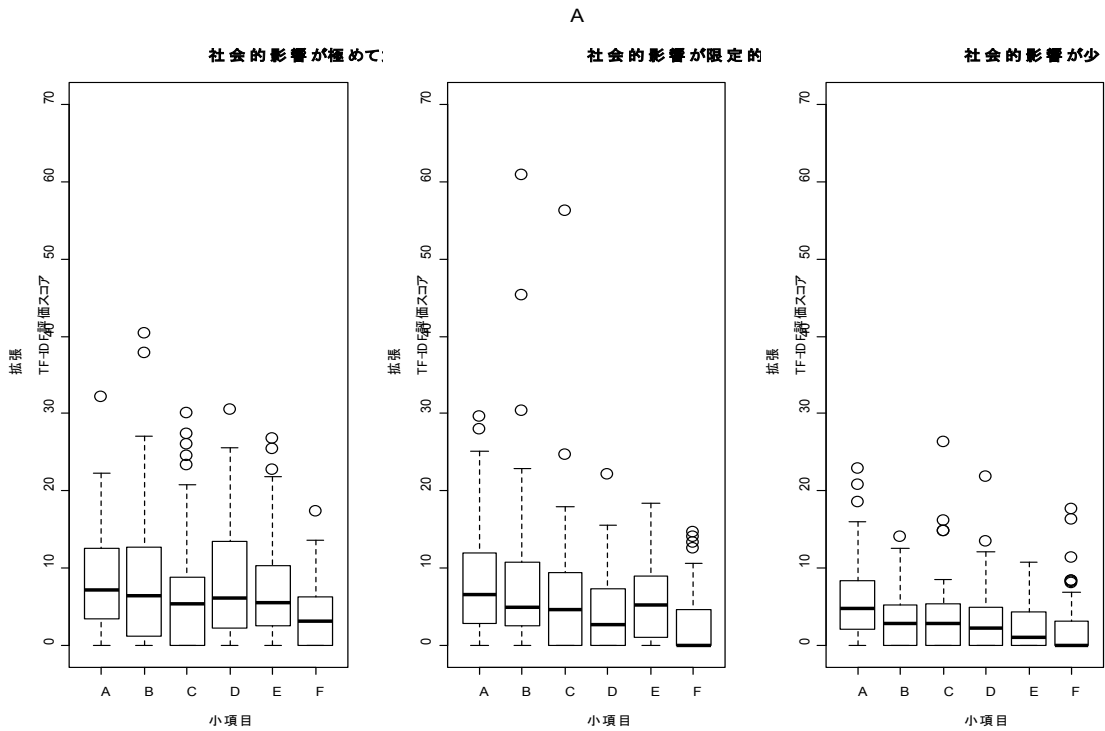


図 5.3 社会的影響カテゴリ別の個別 RFP 評価

5.4 社会的影響カテゴリ別の評価比較

4章で評価対象とした6つの「小項目」を、社会的影響カテゴリ別に「手動評価」した結果をレーダチャートで図 5.4(a)に、また、「拡張 TF-IDF」により評価した結果をレーダチャートで図 5.4(b)に示す。図 5.4 によると、「手動評価」と「拡張 TF-IDF」の評価スコアによる比較でみると、個別の「小項目」についての評価に相違があることが認められる。これらの相違点は、「小項目」の評価を平均値として比較した結果である。しかし、6個の「小項目」それぞれについて、要求される非機能要件の特徴が、社会的影響度合いの差として評価されていることは認められる。

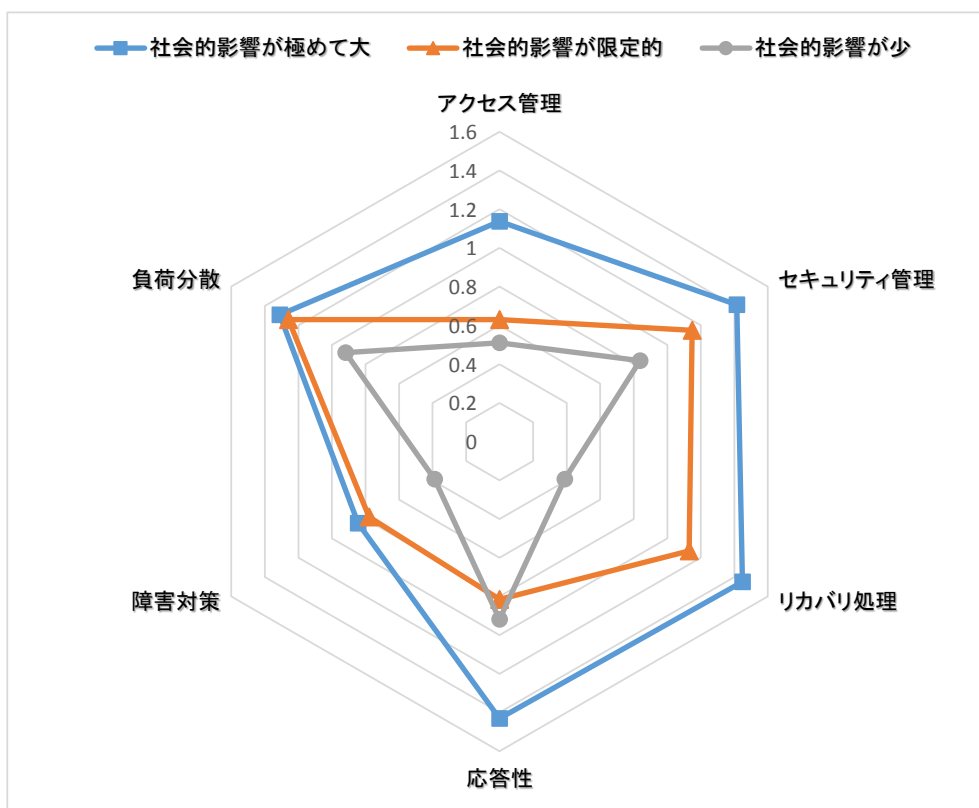


図5.4(a) 手動評価による社会的影響別の小項目評価スコア

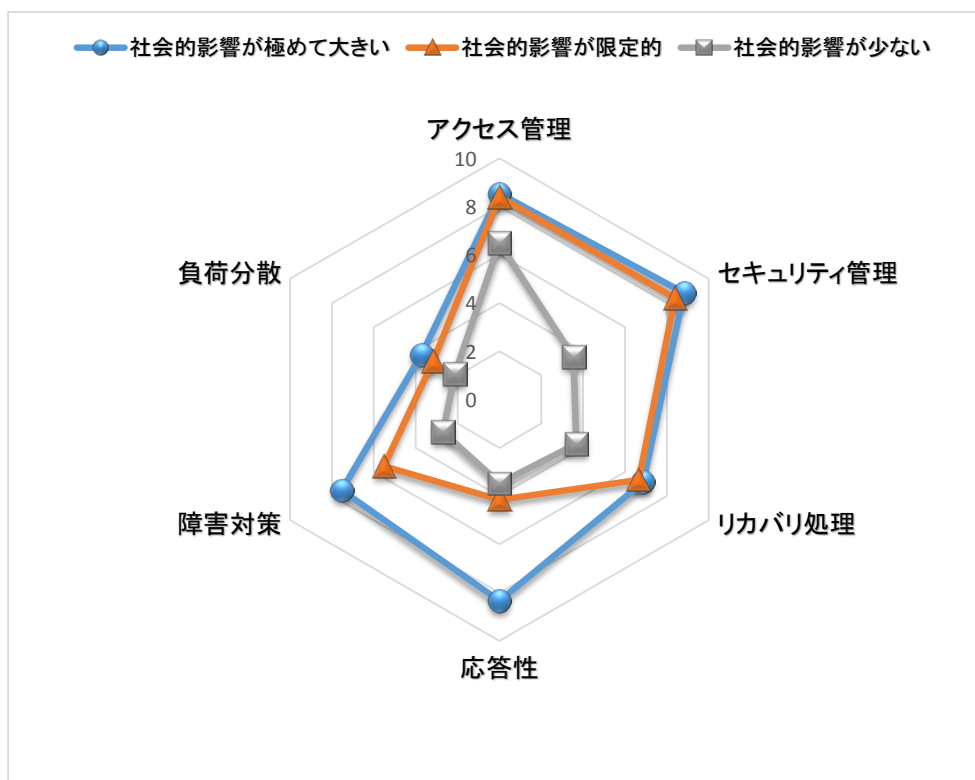


図5.4(b) 社会的影響カテゴリー別の小項目拡張TF-IDF評価スコア

5.5 まとめ

社会的影響カテゴリにより、161件の情報化対象システムであるRFPを3つのカテゴリに分類して評価した結果、「社会的影響が極めて高い」カテゴリのシステムであっても非機能要件である「小項目」については、明確に記述されていないRFPがあることがわかった。また、同じカテゴリのRFPであっても「セキュリティ管理」や「リカバリ処理」については、その評価スコアに大きな差があり、非機能要件である「小項目」の記述内容の明確さに差があることもわかった。これらのことは、「NFR キーワード評価シート」に基づく評価により非機能要件の記述内容の明確さにおいて、明確さが十分でない「小特性」を定量評価することの必要性を示唆していると言える。ただし、社会的影響によるカテゴリ分類は、定量的に明確ではないため、分類された個々のRFPについて、そのカテゴリに属するかどうかについては議論の余地がある。しかし、社会的影響により3つにカテゴリ化されたそれぞれの評価スコアの平均値でみると、「応答性」、「障害対策」および「負荷バランス」において、「社会的影響が極めて高い」システムについては評価が高く、一般的な認識と一致することが認められた。また、社会的影響別の評価結果については、社会的影響が高いほど評価が高いことも認められた。これらのことは、RFPに記述されるべき、どの非機能要件が明確に記述されていないかについて定量的に把握できることを示している。したがって、委託ソフトウェア開発の超上流工程で、「要件定義」をより明確にすることを支援するために「NFR キーワード評価シート」の活用は、有効な手法であると考えられる。

第6章 結論

本論文では、委託ソフトウェア開発の超上流工程でユーザが作成した RFP の非機能要件の記述内容を定量的に評価する評価指標を示し、RFP に記述されている非機能要件の明確化を支援することを目的として、(1)ユーザにとって重要度の高い「保守と運用に関する 55 個の非機能要件」で構成した「非機能要件評価シート」による評価モデル、(2)RFP テキストファイルから抽出し「非機能要件評価シート」の「中項目」に対応させた「NFR キーワード評価シート」を作成し、NFR キーワード出現頻度による機械学習手法を用いた自動評価モデル、(3)機械学習によらない 3 つの手法による NFR キーワードの重みで比較評価し、追加実験として教師なし機械学習の手法であるクラスタリングによる 3 段階評価、最後に(4)RFP を社会的影響により 3 つのカテゴリに分類し、3 つの手法の中で有効であった「拡張 TF-IDF」により評価を行い、「NFR キーワード評価シート」による非機能要件の評価の有効性を検証した。

最初のステップでは、文献およびウェブに公開されている調査結果をもとに「非機能要件評価シート」を作成し、非機能要件である「小項目」による RFP の非機能要件の記述内容の明確さの評価が可能であるかについてケーススタディを行った。ケーススタディでは、ウェブ上に公開されている 29 件の RFP データを「非機能要件評価シート」に基づいて評価した結果を情報化対象ドメインごとに分類して比較評価した。さらに、29 件の RFP 評価結果から、評価の高い 3 個の RFP を評価基準とした RFP の相対比較を行った。RFP の比較評価を行った結果、どの「小項目」(非機能要件)による記述内容が明確でないかについて情報化対象ドメイン毎に定量的に視覚化できることがわかった。また、「非機能要件評価シート」に基づいて 3 人の評価者が評価した一部の「小項目」(非機能要件)について、評価者によるばらつきに有意差があり、今後はより多くの評価者による手動評価のばらつきを抑えることが課題である。

第 2 のステップでは、RFP に記述されている NFR キーワードの出現頻度を説明変数とし、手動評価結果を教師データ(目的変数)としたランダムフォレスト手法による自動評価実験を行った。実験では、ウェブ上に公開されている 70

件の RFP データをテキスト変換した後にテキストマイニングによる NFR キーワード抽出を行った。抽出した NFR キーワードを関連付けた 26 個の「小項目」(非機能要件)について 2 段, 3 段および 5 段の手動段階評価を教師データとしてランダムフォレストによる評価推定実験を行った。ランダムフォレストによる評価推定と手動評価との一致率を比較および κ 統計値により検証した。実験の結果, 単純比較による一致率は, 69.8%であった。また, 26 個の NFR 指標を平均した 3 段階評価の κ 統計値は 0.4063 であり, κ 統計値の判定基準で「中程度の一致」と判定された。これは, 2 段階評価および 5 段階評価との比較において高い一致率であり, 3 段階評価が RFP の NFR キーワードによる自動評価モデルとして, 非機能要件の記述の明確化を支援することに有効であることが分かった。ただし, NFR キーワードの選定や数を変えてモデルのチューニングを行い, ランダムフォレストの精度を高めることが今後の課題である。

第 3 ステップでは, RFP に記述されている NFR キーワードによる評価法として「単純 TF-IDF」, 「拡張 TF-IDF」および「キーワード比」の 3 つの手法を NFR キーワードの重みとして評価実験を行った。ウェブ上に公開されている 161 件の RFP データを収集し, 評価結果を比較した結果は, 「拡張 TF-IDF」手法で評価対象とした 6 つの「小項目」のすべてにおいて, 手動評価との相関係数が高いことが分かった。しかし, 「拡張 TF-IDF」と「手動評価」との比較において, 各「小項目」に外れ値となった RFP があり, NFR キーワードによる評価に文章の前後関係や文脈も含めた手法の改良が今後の課題である。

次に, この実験結果をもとに k-means 法によるクラスタリングの追加実験を行った。6 個の「小項目」についてのクラスタリングと手動評価との一致率は 60.5%であった。一方, クラスタリングの評価尺度であるエントロピー及び純度はそれぞれ 0.737 と 0.606 であり, エントロピーを低く, 純度を高くするアルゴリズムと組み合わせることが今後の課題であることが分かった。

さらに, 「NFR キーワード評価シート」による非機能要件の評価の有効性を確認するために 161 件の RFP を社会的影響による 3 つのカテゴリに分類して, 第 3 ステップで提案した「拡張 TF-IDF」法による評価と手動評価の比較を行った。算出した評価スコアにおいて, 「社会的影響が極めて大きいシステム」で

あっても非機能要件の記述が明確でない「非機能要件」があることが、「社会的影響が限定されるシステム」や「社会的影響が少ないシステム」との比較によって明らかになった。しかし、さらに多くの RFP による実験と検証が課題であることも分かった。

本論文の成果により、超上流工程でユーザが作成した RFP の非機能要件の記述内容を「非機能要件評価シート」および「NFR キーワード評価シート」を用いた提案手法で定量的に評価し、評価指標とすることを示すことができた。今後は、提案手法の精度向上と有効性を高め、非機能要件の明確化を定量評価することにより、委託ソフトウェア開発の超上流工程におけるユーザ要件定義を支援する研究を続けたいと考えている。

謝辞

本研究を進めるにあたり、ご指導、ご協力頂いた皆様に感謝の意を表したいと思えます。本当にありがとうございました。

研究の過程で、適宜懇切なご指導、ご助言と共に、適切なご指摘を賜りました、奈良先端科学技術大学院大学 情報科学研究科 松本 健一 教授 に、心から深く感謝申し上げます。

本研究を進めるにあたり、研究発表での貴重なご指導、ご助言を賜りました、奈良先端科学技術大学院大学 情報科学研究科 藤川 和利 教授に、心から深く感謝申し上げます。

本研究を進めるにあたり、研究の方向性を決める有益なご助言と終始熱心に相談に応じて頂き、論文執筆に際しては丁寧なご指導をいただきました、奈良先端科学技術大学院大学 情報科学研究科 門田 暁人 准教授 に、心から深く感謝申し上げます。

本研究を進めるにあたり、研究発表や会議での貴重なご指導、ご助言を賜りました、奈良先端科学技術大学院大学 情報科学研究科 伊原 彰紀 助教に、心から深く感謝申し上げます。

本研究を進めるにあたり、奈良先端科学技術大学院大学 情報科学研究科 ソフトウェア工学講座の皆様には、研究活動や学生生活におきましてさまざまなご支援を賜りました。心から深く感謝申し上げます。

最後に、精神的にも長く、時にはくじけそうになった研究生生活と本稿執筆に至る道のりにおいて、支え励ましてくれた家族である長男と次男、そして家を離れての学生生活を見守り、支えてくれた 妻 美和子 に心から感謝します。

参考文献

- [1] 情報処理推進機構 ソフトウェア・エンジニアリング・センター, “共通フレーム 2007”, オーム社, 2007.
- [2] 情報処理推進機構 ソフトウェア・エンジニアリング・センター, “非機能要求の見える化と確認の手段を実現する「非機能要件グレード」”, 2010.
- [3] 日本情報システム・ユーザー協会編, “非機能要求仕様定義ガイドライン”, 2008.
- [4] 日本情報システム・ユーザー協会, “ソフトウェア開発管理基準に関する調査報告書(ソフトウェアメトリクス調査)”, 2012.
- [5] Rick Kazman, Mark Klein, Mario Barbacci, Tom Longstaff, Howard Lipson, Jeromy Carriere, “The Architecture Tradeoff Analysis method”, Technical Report CMU/SEI-98-TR-008, ESC-TR-98-008, Carnegie Mellon University, Software Engineering Institute, 1998.
- [6] 経済産業省ソフトウェアメトリクス高度化プロジェクトプロダクト品質メトリクスWG, “システム/ソフトウェア製品の品質要求定義と品質評価のためのメトリクスに関する調査報告書”, 2011.
- [7] 日経ソリューションビジネス編, “システム構築のトラブルを回避するためのITシステム契約締結の手順とポイント”, 日経 BP 社, 2008.
- [8] Bud Porter-Roth(著), 渡部洋子(訳), “RFP 入門 —初めての提案依頼書”, 日経 BP 社, 2004.
- [9] 齊藤康廣, 門田暁人, 松本健一, “ソフトウェア委託開発プロジェクトの超上流工程における非機能要件評価に関する研究”, 奈良先端科学技術大学院大学テクニカルレポート, NAIST-IS-TR2013001, 2013.
- [10] 佐藤知徳, 鈴木俊一, 北澤直幸, 長田晃, 海谷治彦, 海尻賢二, “ソフトウェア要求仕様における品質要求の含有率測定ツールの設計”, 電子情

- 報通信学会技術研究報告(知能ソフトウェア工学 KBSE2007-57), Vol. 107, No.540, 2008, pp.19-24.
- [11] 経済産業省 商務情報政策局 情報処理振興課,情報処理推進機構, “情報システム調達のための技術参照モデル(TRM)平成 22 年度版” 2011.
- [12] 今村誠, 高山泰博, 三上崇志, 岡田康裕, “技術文書からの用語知識自動獲得の検討”, 情報処理学会研究報告, 2007(34), 2007, pp. 25-32.
- [13] 小林雄一郎, 田中省作, 富浦洋一“N-gramを素性とするパターン認識を用いた英語科学論文の質判定”, 情報処理学会研究報告 2012-IFAT-105, 2012, pp.1-6.
- [14] 小林雄一郎, 田中省作, 富浦洋一“メタ談話標識を素性とするパターン認識を用いた英語科学論文の質判定”, 人文科学とコンピュータシンポジウム論文集, 2011, pp.51-58.
- [15] 齊藤康廣, 門田暁人, 松本健一. “Request For Proposal(RFP)における保守・運用要件指標の抽出と評価”, 情報処理学会研究報告 2012-SE-175(1), 2012, pp.1-10.
- [16] 齊藤康廣, 門田彰人, 松本健一, “非機能要件構築のためのデータ”, 奈良先端科学技術大学院大学, <http://se-naist.jp/NFR/> , 2014.
- [17] 金明哲, 村上正勝. “ランダムフォレスト法による文書書き手の同定”, 数理統計, 第 55 巻第 2 号, 2007, pp.255-268.
- [18] 樋口耕一, “KH_Coder2.x レファレンスマニュアル”, 2012, pp.84.
- [19] 福田淳一, “テキストマイニングのシステム開発上流工程適用の試み”, Journal of the Society of Project management Vol. 13, N0.2, 2011.
- [20] Agustin Casamayor, Daniela Godoy, Marcelo Campo“Functional Grouping of Natural Language Requirements for Assistance in Architectural Software Design”, Knowledge-Based Systems Vol.

30, 2012, pp.78-86.

- [21] Agustin Casamayor, Daniela Godoy, Marcelo Campo, “Mining Textual Requirements to Assist Architectural Software Design: A State of the Art Review”, *Artificial Intelligence Rev.* 38,2012, pp.173-191.
- [22] Carlos Huertas, Reyes Juárez-Ramírez “NLARE, “A Natural Language Processing Tool for Automatic Requirements Evaluation”, *Proceeding of International Conference on Cloud and Ubiquitous Computing and Emerging Technologies (CUBE’ 2012)*, September 2012, pp.3-5.
- [23] Dan Port, Allen Nikora, Jane Huffman Hayes, LiGuo Huang, “Text Mining Support for Software Requirements: Traceability Assurance”, *Proceedings of 44th Hawaii International Conference on System Sciences*, 2011, pp.1-11.
- [24] Leonid Kof, “Natural Language Processing for Requirements Engineering: Applicability to Large Requirements Documents”, <http://www4.informatik.tu-muenchen.de/>, 2004.
- [25] Gokhan Gokyer, Semih Cetin, Cevat Sener, Meltem T. Yondem “Non-Functional Requirements to Architectural Concerns: ML and NLP at Crossroads” , *Proceeding of 3rd International Conference on Software Engineering Advances*, 2008, pp.400-406.
- [26] Jane Cleland-Huang, Raffaella Settini, Xuchang Zou, Peter Solc “Automated classification of non-functional requirements” , *Requirements Engineering*12, 2007, pp.103-120.
- [27] Jane Cleland-Huang, Raffaella Settini, Xuchang Zou, Peter Solc “The Detection and Classification of Non-Functional Requirements with Application to Early Aspects” , *Proceeding of*

14th IEEE International Requirements Engineering Conference (RE'06), 2006, pp.36-45.

- [28] Olga Ormandjieva, Ishrar Hussain, Leila Kosseim, “Toward a Text Classification System for the Quality Assessment of Software Requirements Written in Natural Language”, Proceeding of 4th International Workshop on Software Quality Assurance (SOQUA'07), September 2007, pp.39-45.
- [29] Simrandeep Singh Thapar, Hardeep Singh, Karanjeet Singh Kahlon, “Metrics-Based Evaluation of Quality of Non-Functional Specifications”, International Journal of Information Technology and Knowledge Management, Volume 2, No. 1, June 2009, pp.131-134.
- [30] Yasuhiro Saito, Akito Monden, Kenichi Matsumoto, “Evaluation of Non Functional Requirements in a Request For Proposal(RFP) ”, Proceeding of International Workshop on Software Measurement and 7th International Conference on Software Process and Product Measurement (IWSM-MENSURA), Oct. 2012, pp.106 - 111.
- [31] Joseph L. Fleiss, “Measuring Nominal Scale Agreement among Many Rater”, Psychological Bulletin, Vol. 76, No.5, 1971, pp.378-382.
- [32] ローレン・フェルドマン, ジェイムズ・サンガー, 辻井潤一監訳, IBM 東京基礎研究所, “テキストマイニングハンドブック”, 東京電機大学出版局, 2010.
- [33] Gerard Salton, Michael. J. McGill, “Introduction to Modern Information Retrieval”, McGraw-Hill, 1983.
- [34] Clay Palmeira, Rafael Chaves, Hamilton Cavalcante, Eloi Favero, “A Requirements Elicitation and Analysis Aided by Text Mining”,

International Journal of Computer Science and Network Security (IJCSNS), Vol.12, No.6, 2012, pp.122-128.

- [35] Fabrizio Fabbrini, Mario Fusani, Stefania Gnesi, Giuseppe Lami, “An Automatic Quality Evaluation for Natural Language Requirements”, Proceedings of 7th International Workshop on Requirements Engineering: Foundation for Software Quality, 2001, pp.150-164.
- [36] Leonid Kof, “Natural Language Processing: Mature Enough for Requirements Documents Analysis?” , Natural Language Processing and Information Systems Lecture Notes in Computer Science Vol. 3513, 2005, pp.91-102.
- [37] Ishrar Hussain, Leila Kosseim, and Olga Ormandjieva, “Using Linguistic Knowledge to Classify Non-functional Requirements in SRS documents”, Natural Language and Information Systems Lecture Notes in Computer Science Vol. 5039, 2008, pp.287-298.
- [38] Anna Perini, Angelo Susi, Paolo Avesani, “A Machine Learning Approach to Software Requirements Prioritization”, IEEE Transactions on Software Engineering Digital Object Identifier 10.1109/TSE.2012.52, 2012.
- [39] 金明哲, “フリーソフトによるデータ解析・マイニング 第 60 回 統計的テキスト解析 (5) ～統計法則と指標～”, 統計情報研究開発センター ESTRELA (No.172), July 2008.
- [40] 新納浩幸, “R で学ぶクラスター解析”, オーム社, November 2009.
- [41] 長谷川亮, 北村元博, 海谷治彦, 佐伯元司, “要求分析のためのドメインオントロジ構築支援”, 電子情報通信学会技術研究報告. SS, ソフトウェアサイエンス 107(176), July 2007, pp.53-58.
- [42] 情報処理機構 ソフトウェア・エンジニアリング・センター, “非機能要求グレ

ード利用ガイド[解説編]”, April 2010.

- [43] 齊藤康廣, 門田暁人, 松本健一, “RFPにおける機械学習による非機能要件の評価”, 研究報告ソフトウェア工学(SE), 2013-SE-179(5), (2013-03-04), pp.1-7.
- [44] René Witte¹, Qiangqiang Li¹, Yonggang Zhang, and Juergen Rilling, “Ontological Text Mining of Software Documents”, Springer-Verlag Berlin Heidelberg 2007 NLDB 2007, LNCS 4592, (2007), pp.168-180.
- [45] 情報処理機構 ソフトウェア・エンジニアリング・センター, “IPA/SEC が提案する ソフトウェア品質指標の解説 ～品質作り込みガイド(ESQR)の概要解説～”, www.ipa.go.jp/files/000030337.pdf, January 2013.
- [46] Capers Jones 著, 富野壽・小坂恭一監訳, “ソフトウェア開発の定量化手法 第3版”, 共立出版株式会社, July 2010.
- [47] Stephen H. Kan 著, 古山恒夫・富野壽 監訳, “ソフトウェア品質工学の尺度とモデル”, 共立出版株式会社, November 2004.
- [48] 酒巻弘晃, 横山真一郎, 劉功義, 石井信明, 田村智幸, 牛嶋一朗, 加藤俊, 笠井直貴, 木野泰伸, “RFP の評価と品質向上に関する研究”, プロジェクトマネジメント学会予稿集, 2009 春季, pp. 374-379.
- [49] 山本修一郎, “～ゴール指向による!!～システム要求管理技法”, 株式会社ソフトウェア・リサーチ・センター, May 2007.
- [50] エレン・ゴッテスディナー著, 三島邦彦, 前田卓雄, 宗雅彦 監訳, 成田光彰 訳, “要求開発ワークショップの進め方ーユーザ要求を引き出すファシリテーションー”, 日経 BP 社, July 2007.

付録 A. 「NFR キーワード評価シート」 1/6

大項目	中項目	小項目	NFRキーワード		
運用開始の準備要件	FAT特性	運用テスト計画	APIテスト システムテスト計画書 システムテスト計画 システムテスト実施手順書 システムテスト システム総合テスト 総合テスト計画書 データ移行計画書 データ移行仕様書 データ移行仕様	結合テスト計画 テスト実施計画書 テスト実施スケジュール テスト計画書 テスト計画 テストシナリオ 受入テスト計画 運用テスト計画書 運用テスト手順書 運用テスト仕様書	セキュリティテスト計画書 移行実施計画 移行計画 移行手順書 システム移行計画書 システム移行計画 システム移行 運用計画書 試験計画書 接続テスト計画書
		運用テスト実施	総合テスト仕様書 テスト仕様書 テスト項目表 テスト項目 テスト方法 テスト範囲 テスト環境 テスト実施	運用テスト期間 データ移行期間 テスト実施期間 稼働準備 移行作業マニュアル 稼働移行 移行手順書 移行環境	結合テスト支援 データ移行 移行データ データ移行作業 システム移行ツール 移行処理 移行期間
	運用開始条件	結合テスト	ストレステスト 再始動テスト システムレベルテスト システムテスト 性能テスト データ移行テスト	信頼性テスト 接続テスト 障害対策テスト 操作性テスト トラブルテスト	インストールテスト 信頼性テスト 性能評価テスト セキュリティテスト 結合テスト
		総合テスト	システム総合テスト 総合テスト システム環境稼働確認テスト 本番環境稼働テスト システム本番稼働 システム本稼働 システム稼働開始 ユーザ運用テスト 受け入れテスト	システム稼働 システム移行 試験運用 テスト実施 稼働移行 テスト運用 稼働確認テスト 受入テスト 受入テスト実施	稼働テスト 運用テスト実施 テスト稼働 導入テスト 稼働テスト 運用テスト ユーザテスト
テスト結果評価		総合テスト完了報告書 テスト報告書 データ検証 システム検証	検証システム 運用前提条件 総合テスト終了 システム稼働立会報告書	システム稼働立ち会い 稼働日 本番稼働	
システム運用の評価要件	操作容易特性	オペレーション	操作性 操作マニュアル システム操作方法 操作方法 操作手順書 システム操作 操作説明書	操作説明 操作手順 システム運用手順書 運用手順書 運用手順 操作環境 入力ミス	操作ガイド 操作ガイド機能 オンラインマニュアル 画面遷移
		稼働品質特性	応答性	平均読み出し遅延 転送応答性 応答性 ハードディスク応答性能 画面レスポンス オンライン応答時間 平均処理応答 スループット レスポンスタイム 秒程度	最大スループット 安定的レスポンス 最小レスポンス 最小応答 ネットワーク応答 VPNスループット 平均応答 応答時間 秒以内 ms以下
	システム性能	性能監視機能 処理速度 システム性能 MPU使用率 CPU使用率 性能目標値 オンライン処理性能 性能 CPU	処理性能 アクセス速度 CPU性能 CPU負荷率 性能評価 ネットワーク使用率 %以下 CPU SPEC	アクセス性能向上 ディスクIO負荷率 性能耐久性 サーバ性能 同時接続数 性能管理機能 演算性能 中央演算処理装置	
	負荷バランス	負荷率 負荷分散性能 負荷分散 負荷許容範囲 最小サーバ負荷 ネットワーク負荷 最大負荷時 負荷予測	負荷分散方式 負荷分散構成 負荷低減 負荷運用 サーバ負荷 負荷分散装置 負荷監視 ロードバランシング	負荷分散機能 負荷計測 動的負荷分散機能 運用負荷 ロードバランス ピーク時 CPU負荷	
稼働品質	稼働実績 安定稼働実績 システム正常性 トラザクションデータ アクセス頻度 通信速度 稼働安定性 パフォーマンス管理 365日安定稼働 稼働状況 運用スケジュール 稼働期間 24時間365日	トラザクション数 トラフィック量 業務稼働率 ダウンタイム トラザクション管理 アクセス量 平均稼働率 サーバ稼働率 稼働状況 安定システム 稼働実績 %以上 サーバ稼働率	稼働率 システム稼働率 年間稼働率 連続稼働 安定運用 連続運転 システム停止許容 正常稼働 システム稼働状況 システム稼働情報 ネットワーク管理状況 %以下		

「NFR キーワード評価シート」 2/6

大項目	中項目	小項目	NFRキーワード
運用監視要件	障害検知特性	システム異常検知	不具合発生 停電信号 状態監視機能 監視機能 異常発生通知 自動運転監視 故障検知 動作監視 トラブル発生 自己診断機能 モニタリング システム障害 接続状況監視 状態監視 監視ソフト 異常動作 検知システム 事故発生 統合監視システム 状態情報 システムダウン システム稼働監視 システム監視 システム停止 上限値警告 監視システム 異常メッセージ イベント発生 監視内容 システムダウン 運用監視サーバ システム稼働管理サーバ システムエラー 異常
		障害要因	障害要因 障害発生場所 障害発生検知 障害内容 障害情報 障害監視機能 障害の内容 障害切り分け 障害の原因分析 障害システム 障害の分析 プラットフォーム障害 ハードウェア障害 サーバ監視ソフトウェア ハードディスク障害 電源ユニット障害 プリンタ監視 障害発生通知 障害発生時点 障害発生箇所 障害通知機能 障害原因 障害監視 障害時 ハード障害発生 障害理由 ネットワーク障害 原因切り分け ディスク障害 サーバ異常通知機能 ディスク障害用 サーバ故障 ハードウェア障害発生 ネットワークダウン サーバ異常 障害発生状況 障害発生時刻 障害発生 障害状況 障害検知 障害箇所 システム障害1次切り分け ソフトウェア障害 障害の切り分け サーバ障害検出 アプリケーション障害 サーバ障害 電源障害 機器障害 ネットワーク障害 障害検知 データベース機能停止
	システム監視特性	運用管理	稼働状態 運用維持管理 運用支援 運用管理システム 運用管理ソフトウェア 運用管理ソフト 運用権限 運用管理 運用ルール 運用計画 運用管理機能 運用監視ソフト 運用監視 運用マニュアル 運用状態 運用権限 運用管理ツール 運転状況 運用スケジュール 稼働状況 運用状況 運用管理要件 正常運用 運用時間 運用管理マニュアル 運用保守 運用・管理 運用時間帯
		システム管理	ネットワーク監視装置 ネットワーク監視機能 ネットワーク監視機能 ネットワーク監視 ネットワークシステム管理 ネットワーク管理 システム稼働状況 システム運用支援機能 システム監視手順 運用監視システム システム運用機能 運用監視機能 システム監視用 システム監視機能 システム稼働管理 システム監視 システム管理 システム管理機能 システム運用管理機能 システム運用管理 システム運用管理担当者 運用管理サーバ 稼働監視 ファイル監視 リモート監視 サーバ運用監視 サーバ運用管理 サーバ管理機能 サーバ管理 サーバ監視機能 内部機器監視サーバ システム管理者 運用管理障害監視 監視ソフトウェア 遠隔監視
セキュリティ対策特性	不正アクセス		漏えい防止 不正操作対策機能 不正アクセス対策機器 不正アクセス 侵入対策 ファイルアクセス制御 脆弱性対策 脆弱性情報 脆弱性 不正アクセス防止 不正アクセス対策 不正侵入防止 侵入監視 アクセス制御機構 脆弱性 不正アクセス防御対策 不正アクセス検知 不正侵入 アクセス制御 アクセス権管理システム アクセス権制御 アクセス権管理機能 アクセス権 権限管理 管理権限 アクセス制限 アクセスコントロール機能 権限区分 アクセス権限 操作権限 権限レベル ユーザ権限 アクセスポリシー アクセスコントロール 権限設定 管理者権限 制限ユーザ 業務権限 ユーザ数 ユーザ権限 アカウント管理 利用権限 利用者権限登録機能 アクセス管理基盤 利用者権限管理 利用者数コントロール 権限要件 権限レベル
		アクセス権限	
	セキュリティ管理レベル		情報セキュリティ対策統一 情報セキュリティ対策基準 情報セキュリティレベル 情報セキュリティ実施手順 情報セキュリティ管理システム 監査業務 監査情報 監査証跡 セキュリティに関する研修 セキュリティ確保 セキュリティ運用 セキュリティポリシー管理機能 セキュリティポリシー 情報セキュリティポリシー ローカルセキュリティポリシー セキュリティポリシー セキュリティの確保 セキュリティ確保 セキュリティ管理要領 セキュリティの維持 セキュリティ管理システム セキュリティ管理計画書 セキュリティ管理機能 セキュリティ監査 セキュリティ対策レベル セキュリティレベル セキュリティリスク分析機能 インシデント管理 セキュリティリスク セキュリティ教育 セキュリティ診断 セキュリティ管理

「NFR キーワード評価シート」 3/6

大項目	中項目	小項目	NFRキーワード		
運用監視要件	セキュリティ対策特性	セキュリティ対応	セキュリティワイヤー セキュリティホール セキュリティパッチ セキュリティシステム スパムブロック機能 サイバー攻撃 シンククライアント方式 セキュリティロック シンククライアント方式 シンククライアント端末 VPN接続装置 セキュリティパッチ適用 セキュリティ侵害監視 セキュリティ対策レベル セキュリティテスト セキュリティ機能	セキュリティチェック セキュリティモニタリング セキュリティホール対策 セキュリティパッチ管理 セキュリティゾーン セキュリティシステム シヤドウパスワード機能 VPN接続機能 検疫ネットワーク セキュリティ実施手順 セキュリティソフトウェア パッチ管理システム セキュリティ対策ソフト セキュリティレベル ネットワークセキュリティ セキュリティソフト	仮想化セキュリティサーバ セキュリティ設定 データセキュリティ ネットワークセキュリティ ネットワークセキュリティシステム セキュリティ監視 セキュリティ情報 セキュリティ対策状況 セキュリティ対応 セキュリティ対策 セキュリティ侵害監視 セキュリティ侵害 セキュリティレベル 情報セキュリティ対策 情報セキュリティ セキュリティ脅威
		認証機能	利用者認証機能 本人認証 ユーザID 利用者認証 操作者認証 クライアント認証 生体認証 不正監視 認証方式 認証基盤 認証システム 認証VLAN ログイン認証 主体認証方式 認証方式 SSLクライアント認証機能	ユーザ認証システム ユーザ認証 職員認証 利用者認証機構 ユーザ認証システム ログオン認証 認証機能 認証管理 統合認証システム ネットワーク認証サーバ ネットワーク認証 認証基盤システム 認証ソフトウェア 認証クライアントソフトウェア 認証対象者数	ユーザ認証基盤 ユーザ認証サービス システムログイン 不正ユーザ 認証情報 主体認証情報 静脈認証 端末認証 ネットワーク認証 MAC認証 サーバ認証 認証サーバ 認証データ 端末認証 統合認証データサーバ
		パスワード管理	ワンタイムパスワード パスワード認証 パスワード管理 パスワード情報 パスワードポリシー パスワードロック 利用者パスワード 旧パスワード パスワードの変更機能 パスワードの有効期限	パスワード入力 パスワード一元管理 初期設定パスワード ログイン制御 ロックアウト処理 ログオン制御 利用者のパスワード パスワード有効日数 強制パスワード変更 パスワード登録	ログインパスワード ユーザ認証パスワード パスワード設定 パスワード変更 認証管理 パスワード管理機能 パスワード入力 初期パスワード パスワードの暗号化
		ウイルス対策	不正プログラム ウイルス検出 ウイルス感染 パターンファイル ウイルスパターンファイル ウイルスチェックソフト ウイルスチェック ウイルスパターンファイル 検疫ネットワーク機能 検疫サーバ 検疫対象 検疫ネットワーク 検疫エージェント 検疫機能 検疫セグメント ウイルス除去	ウイルスソフトウェア ウイルスソフト ウイルススキャンエンジン ウイルス検知 ウイルスチェック ウイルス定義ファイル アンチウイルスソフトウェア アンチウイルス機能 ウイルス対策ソフトウェア ウイルス対策システム ウイルス対策ソフト ウイルス対策 ウイルス対策ソフトWebサーバ ウイルス対策サーバ ウイルス管理システム ウイルス管理ソフト	ウイルス監視 攻撃パターン ウイルス検知報告 ウイルスソフト ウイルスチェックサーバ ウイルス ウイルス定義 ウイルスパターン更新 ウイルスチェックソフトウェア ウイルススキャン機能 コンピュータウイルス対策 ウイルス対策機能 ウイルスチェック対策 ウイルス監視 ウイルス防御
		情報漏洩対策	情報漏洩対策機器 情報漏洩対策エージェント 情報漏えい防止 改竄防止 データの漏洩	情報漏洩対策ソフトウェア 情報漏洩対策 情報漏えい 改ざん防止	情報漏洩対策サーバ 情報漏洩 データ漏洩
		暗号処理	HDD復号化 復号化 暗号化 ハッシュ化 ssh暗号化通信 SSL暗号化通信 暗号化通信 暗号方式	通信暗号化 公開鍵基盤 鍵管理定義書 鍵管理 秘密鍵 公開キー交換 暗号化キー 暗号化通信路	暗号化対策 ファイル暗号ソフトウェア 暗号化通信機能 暗号化強度 鍵管理定義書 暗号通信モード 暗号化ファイルサーバ機能
障害対策要件	冗長化特性	冗長化	二重化構成 電源二重化 多重化対策 多重化 冗長化機構 冗長構成 冗長化構成 冗長化 仮想化技術	冗長性 冗長化電源 冗長化度合い デュアルディスプレイ構成 冗長電源 バックアップ電源 リダンダント機能 ホットスベア	ディスクアレキシシステム ディスクアレイ コールドスタンバイ ホットスベアディスク コールドスタンバイ方式 分散システム 仮想化管理サーバ 仮想化サーバ
		RAID構成	ハードウェアRAID-1構成 メモリミラーリング機能 ミラー構成 ミラーリング機能 ミラー化 ミラーリング ハードウェアRAID クラスタリング	RAIDコントローラ RAID0+1構成 RAIDレベル RAIDグループ RAID RAID機能 RAID構成 クラスタシステム	RAIDレベル RAID01構成 RAID5ストレージ RAID5構成 RAID5 クラスタ構成 RAID

「NFR キーワード評価シート」 4/6

大項目	中項目	小項目	NFRキーワード		
障害対策要件	データバックアップ特性	バックアップ管理	日中バックアップ	フルバックアップ	データバックアップ機能
			定期バックアップ	一括バックアップ	データバックアップ方式
		日次バックアップ	フルバックアップイメージ	バックアップ対象データ	
		定期的バックアップ成功率	フルバックアップ運用	バックアップイメージ	
		自動バックアップ	フルバックアップ	バックアップデータ	
		システムバックアップ	暗号化バックアップ機能	バックアップ方式	
		バックアップデータ	各種ログバックアップ	データバックアップ	
		データバックアップ	データバックアップスケジュール	データ回避保存	
		差分ブロック	バックアップスケジュール	データ保護	
		差分データ	バックアップデータ項目	バックアップ管理	
差分バックアップ	バックアップ期間	バックアップ仕様			
世代保管	バックアップ速度	バックアップ運用			
差起動	バックアップ手順	データのバックアップ			
差分データバックアップ	データバックアップ方法	バックアップ対象			
バックアップを自動	定期的バックアップ	世代管理			
バックアップシステム	バックアップ転送	バックアップソフトウェア	専用バックアップサーバ		
	バックアップ実行	バックアップソフト	バックアップ専用サーバ		
バックアップ機能	バックアップシステム	バックアップ装置			
バックアップツール	バックアップソフト	バックアップ媒体			
バックアップ統合基盤	バックアップ用	バックアップ電源			
バックアップ基盤	バックアップ機構	バックアップ用サーバ			
バックアップ処理	データバックアップ用機器	バックアップサーバ			
障害予防特性	障害対策	耐障害性	システム障害対策	障害対応	
			障害予防	障害時	
		障害対応マニュアル	故障対策	障害調査	
		障害対策マニュアル	障害未然防止	停電対策	
		障害切り分けマニュアル	雷対策	ホットスワップ	
		障害対応方法	障害検知ソフトウェア	フェイルオーバー機能	
		障害対策	緊急対応機能		
		停電装置			
		システム障害対応			
		障害管理	障害履歴	計画停電対応	障害再発防止
障害票	障害発生頻度		障害管理方法		
障害手順書	障害事後対策	障害管理			
障害対応結果報告書	障害管理機能	障害対応計画			
障害再発限度	障害防止管理	電源管理ソフトウェア			
障害管理計画	障害発生限度	障害管理サーバ			
障害管理台帳					
停止処理	冗長稼働	停止期間	計画停止	自動シャットダウン用サーバ機能	
		停電処理	停止システム	自動シャットダウン	
		停止稼働	自動システム停止機能	リモートシャットダウン	
		業務停止	計画停電対応	シャットダウン	
		停止処理	縮退稼働	停電状態	
		非常時緊急運用	保守停止	縮退	
		障害連絡	通常稼働	正常稼働	
		シャットダウン方式	緊急停止	マルウェア対策	
		システムクローズ	シャットダウン	システムクローズ処理	
		強制停止	停止順	正常に停止	
回復方法	自動停止	正常なシステム停止			
障害復旧特性	リカバリ処理	復旧方法	システム復旧	回復レベル	
		復旧対応	システム回復	回復内容	
		復旧対策	リカバリプロセス	回復作業	
		復旧状況	リカバリ処理	障害回復	
		復旧作業	セッションリカバリ機能	障害回復	
		復旧作業対応	データリカバリ処理	回復手順	
		障害復旧	データリカバリ	システム障害回復	
		復旧対応作業	リカバリーディスク	システム回復作業	
		復旧処置	自動リカバリ	データ復旧操作	
		復旧機能	回復処理	データ復旧	
フェイルオーバー	リカバリー	迅速な復旧			
復旧措置	復旧時間	復旧期限			
バッチ処理	プログラム修正バッチ	バッチ適用	動作確認手順書		
	バッチ適用管理機能	バッチファイル適用状況	動作確認		
バッチプログラム	プログラム修正バッチ適用システム起動				
バッチ管理	バッチ対応	バッチ			
修正バッチ	バッチファイル				
災害対策要件	災害対策特性	地震対策	耐震対策	耐震措置	耐震性能
			耐震基準	耐震設備	
		災害発生	天災地変	災害対策機能	
		災害対策サービス	災害対策	災害時	
規模災害	延焼防止対策				
保守生産性要件	問題点把握及び修正分析	運用ログ	操作日時ログ	アクセスログ機能	ログ収集
			操作ログ	アクセス履歴	バッチ処理実施ログ
		操作証跡	アクセスログ	ロギング機能	
		オペレーション記録	アクセス情報	ログ情報	
		操作履歴	履歴情報	運用ログ保存用領域	
		操作者名ログ	アクセスログデータ	ログ記録	
		アクセス記録	アクセス数	ログデータ	
		アクセスリスト	データ収集機能	履歴記録機能	
		データベースアクセスログ	ログ収集機能		
		システムログ	システムログ情報	出力ログ	実行ログ
システム監視記録	変更ログ		エラーログ		
システム操作ログ	業務ログ	イベントログ			
トランザクション記録	システムログ	セキュリティログ			
システム監視記録	イベントログ形式	故障記録			
監査ログ	更新ログ	変更履歴			

「NFR キーワード評価シート」 5/6

大項目	中項目	小項目	NFRキーワード			
保守生産性要件	問題点把握及び修正分析	ログ管理・解析	履歴管理機能 履歴管理システム 世代管理 アクセスログバックアップ 履歴管理 更新ログ管理機能 故障履歴管理	送信ログ管理 ログ監視 ログ管理 アクセスログ管理 利用履歴管理機能 ログ解析機能	送信ログ管理 ログ監視 ログ管理 アクセスログ管理 利用履歴管理機能 ログ解析機能	
	保守容易性	操作マニュアル	端末操作マニュアル 操作マニュアル 利用者マニュアル 操作手順書	オンライン操作マニュアル オンラインマニュアル 操作マニュアル作成 各種操作マニュアル	システム管理者向け操作マニュアル システム操作マニュアル オンラインマニュアル 各種操作マニュアル	
		運用管理マニュアル	運用マニュアル 業務運用マニュアル 稼働運用マニュアル 運用マニュアル作成 運用手順書	運用マニュアル 業務運用マニュアル作成 運用管理マニュアル ユーザマニュアル 業務運用マニュアル	管理運用マニュアル システム運用手順書 システム運用マニュアル	
		システム管理マニュアル	システム管理マニュアル バックアップマニュアル データベース定義書 データベース仕様書 ハードウェアマニュアル	ソフトウェア仕様書 システムインストール手順書 データベース管理マニュアル データベース仕様書 インターフェース定義書	システム管理者向けマニュアル システム管理者マニュアル システム利用者マニュアル システム管理者作業マニュアル システム管理者向けマニュアル	
	保守マニュアル	運用保守マニュアル	ネットワーク機器設定報告 サポートデスク運用マニユ. システム保守手順書 各種障害対応マニュアル	サポートデスク運用マニユ 運用保守向けマニュアル 完成図書 システムドキュメント	システム仕様書 システム取扱説明書 システム機能仕様書 システム機能定義書 取扱説明書	
		構成管理管理	データベース更新履歴作成 ハードウェア構成管理機能 ハードウェア構成情報管理 ネットワーク構成情報管理 ソフトウェア構成情報管理 ソフトウェア構成管理機能 ソフトウェア構成管理 ソフトウェア構成情報	構成変更履歴 構成管理機能 構成情報管理 ソフトウェア構成情報管理 構成管理ファイル ハードウェア構成情報管理 ソフトウェア構成管理 機器構成管理	システム構成管理 構成機器管理 構成管理 構成情報収集 構成変更 構成管理業務	
		バージョンアップ	時刻管理 バージョン管理機能 バージョン管理 バージョンアップ作業 システムバージョンアップ リソース管理機能	バージョン管理ツール バージョンレベル管理 バージョンアッププログラム ドキュメント管理 ソフトウェアバージョン	バージョン管理システム バージョンアップ情報 バージョンアップ ドキュメント管理業務 システムバージョンアップ	
	システム構成	システム構成	システム拡張 ハードウェア構成 ハード構成 データ構造 ソフトウェア構成	拡張容易性 システム構成図 ネットワーク構成図 ネットワークシステム構成 ネットワーク構成 ソフトウェア構成図	システム構成 サーバ構成 ディスク構成 拡張性 ネットワーク構成	
		パッケージ	パッケージ製品 パッケージツール パッケージシステム	パッケージソフト パッケージ仕様書 パッケージソフトウェア	ソフトウェアパッケージ パッケージプログラム カスタマイズ量	
		モジュール性	モジュール構成 ソフトウェア構造 システム基本ソフトウェア システムソフトウェア オペレーティングシステム ソフトウェアモジュール オペレーティングシステム	サーバソフトウェア システムファイル データベースソフト データベース管理ソフトウェア システム基本ソフトウェア データ構造 システムデータ	リソース管理機能 システム開発言語 開発言語 FORTRANコンパイラ Cコンパイラ プログラム開発言語 システム開発言語	
	業務運用と利用者支援要件	サービス特性	サービス体制	窓口サービス ヘルプデスク体制 ヘルプデスク業務 運用サポート ヘルプデスク ヘルプデスク情報 ヘルプデスク業務	サポートデスク業務 サポートデスク運営 サポートデスク システムサポート 保守サポート 初期サポート サポート内容	サービス内容 サポート体制 サポートデスク 運用支援 保守支援体制 保守体制 保守サービス体制
			サービス内容	保守点検 保守手順書 保守計画書 保守計画 操作支援 操作指導 支援業務	保守内容 保守設計 保守作業 保守対象 保守手順 保守業務 保守期間	保守サービス メンテナンス作業 定常保守作業 システム保守運用 ネットワーク保守 データ保守 電子化情報提供システム
			サービスレベル	目標値サービスレベル 保守作業履歴 保守管理 サービスレベル項目 サービスレベル管理 サービスレベル設定	サービスレベル協定 サービスレベル サービスレベル基準値 サービス仕様書 サービスレベル管理 サービスレベル協定	SLA定義書 SLA契約 SLA契約内容 保守管理業務 サービス稼働率
			保守契約条件	ライセンス契約	使用権 使用許諾ライセンス ライセンス数 ライセンス契約書 ソフトウェアライセンス管理 ソフトウェアライセンス管理 クライアントライセンス License方式	利用許諾ライセンス 使用許諾契約 使用許諾 使用許諾証書 使用許諾ソフトウェア 現有ライセンス ライセンス状況 ソフトライセンス
		同時使用ユーザ数	最大同時接続数 同時利用ユーザ数 同時実行数 同時アクセスユーザ 同時接続ユーザ数	同時ユーザ数 同時アクセス 必要ライセンス数 ライセンス数 同時利用ライセンス数	同時接続 同時アクセス数 追加ライセンス	

「NFR キーワード評価シート」 6/6

大項目	中項目	小項目	NFRキーワード
		保守タイプ	予防保守計画 定期保守作業 定期保守作業日 定期的保守 運用保守業務 リモート保守 オンサイト保守 ソフトウェア保守 システム保守運用 システムメンテナンス アプリケーション保守 予防保守管理 定期保守 定期的保守 定期保守 定期メンテナンス バックアップシステム保守 システム保守 システム保守業務 システム保守体制 システム運用業務 予防保守 定期保守点検 定期点検 遠隔保守 ハードウェア保守 システム保守業務 システムメンテナンス作業 オンサイト保守 セキュリティ保守
		知的財産権	著作物 著作権法 著作権 所有権 機密保持 著作人格権 著作権侵害 知的財産権 権利侵害 著作者 著作権者 知的財産 権利帰属 瑕疵担保責任 保守業務委託契約 秘密保持契約 保守契約
	障害対応特性	障害対応	問い合わせ対応サービス 非常時対応 障害対応 故障対応 365日障害対応 保守障害対応 障害問合せ 故障対応状況 ハードウェア障害対応 不具合対応 障害対応体制 故障対応業務 365日保守対応 障害対応記録 運用保守業務報告書 保守作業報告 保守報告書
	導入教育特性	研修方式	導入研修 操作説明会 操作研修会 操作訓練等導入スケジュール 集合教育 操作説明 集合研修受講 オンライン研修 操作教育用 操作教育 端末操作研修 操作研修場所 操作研修 集合研修 システム操作研修 システム管理研修 システム運用研修 ネットワーク管理教育
		研修内容	操作研修コース 教育研修内容 操作教育 研修方法 研修受講 研修支援業務 研修環境 教育研修環境 研修コース 教育カリキュラム作成 講習会 研修実施 研修項目 研修開催 教育研修 教育訓練 研修カリキュラム 研修内容 研修会場 研修場所 システム導入研修 システム操作研修サービス 教育研修方法 運用管理研修 運用研修 研修プログラム 定期研修 研修回数 研修日程 研修開始日
		研修機材	研修用テキスト 研修用サブテキスト 端末操作研修用テキスト 操作教育用資料 研修マニュアル 研修テキスト 研修用データ 研修用クライアント機 研修教材 研修用資料 教育研修用教材コンテンツ 教育用資料 研修用システム 研修用 研修機材 研修資料 研修テキスト操作研修用 操作研修用マニュアル 操作研修マニュアル 操作教育用資料 端末操作研修用テキスト
		研修計画	研修計画書 研修計画 研修環境構築計画 教育研修体制 教育訓練計画 教育計画書 研修環境構築 教育研修作業 教育訓練 教育訓練計画 教育計画書 教育研修計画 教育研修
		研修対象	利用者向け研修 利用者研修 担当者向け研修 職員研修計画 受講対象 研修対象人数 研修受講者数 管理者研修 ユーザ教育 利用者研修 新人研修 職員研修 受講者 研修対象職員数 研修受講者 運用管理者研修 利用者教育 職員向け研修 システム管理者研修 受講対象者 個別研修 研修対象者 教育研修対象 保守運用者研修 システム管理者向け研修会

付録 B. 161 件の RFP による「NFR キーワード評価シート」の評価対象「小項目」

応答性 語数:33	応答性 最小応答 平均応答 最小レスポンス 応答速度確保 安定的レスポンス 転送応答性 端末レスポンス ネットワーク応答 画面レスポンス ハードディスク応答性能 平均読み出し遅延	システム応答速度 スループット スループット目標値 最大スループット VPNスループット 応答時間 オンライン応答時間 レスポンスタイム レスポンス目標値 ターンアラウンド 秒程度 秒以内	秒以下 タイムラグ 平均処理応答 フォワーディングレート メモリ使用率 主記憶容量 ハードディスク容量 ネットワーク転送容量 データ量
リカバリ処理 語数:53	復旧方法 復旧対応 復旧対策 復旧状況 復旧作業 復旧作業対応 障害復旧 復旧対応作業 復旧処置 復旧機能 復旧措置 故障復旧 障害回復用 障害回復 障害復旧対応 回復処理 障害リカバリ手順 回復手順	迅速な復旧 回復作業 データリカバリ処理 データリカバリ データ復旧操作 データ復旧 データ復元 データ復元ポイント システム復旧 システム回復 システム障害回復 システム回復作業 システムリカバリ リカバリプロセス リカバリ処理 ダウンリカバリ リカバリ運用 セッションリカバリ機能	リカバリー 起動確認 端末リカバリ リカバリーディスク 自動リカバリ リカバリディスク リカバリディスク作成 リカバリ対象機器 リカバリログ 障害回復用ログファイル 障害回復テスト 復旧時間 回復時間 回復レベル 回復内容 復旧期限
アクセス管理 語数:71	不正アクセス 不正アクセスを検知する機能 不正アクセスを検知できる機能 不正なアクセス 不正アクセスを発見 不正侵入 不正侵入検知 不正侵入の検知 不正侵入障害発生 不正アクセス監視状況 不正ユーザ 侵入監視 不正アクセス検知 不正ログイン 不正コード 不正アクセス対策 不正アクセス防御対策 不正侵入監視手順 不正侵入障害対応手順 不正接続検知 不正侵入検知パターン適用 不正アクセス防止 不正アクセスを防止 不正アクセスを制御 不正アクセスの監視・記録 侵入対策 不正侵入防止 不正アクセスの解析 不正侵入防止機能 不正侵入保護装置 なりすまし防止 不正アクセス報告 不正アクセス状況	権限が設定されたユーザ 更新権限 利用者ごとの権限 利用者の権限 閲覧権限 ユーザ権限 アクセス権 業務権限 権限要件 利用権限 管理者権限 ファイルアクセス権 制限ユーザ アクセス認証 権限変更 変更権限 アクセス権限機能追加 アクセス権更新 権限変更発生 利用者権限登録機能 アクセス権限設定 権限設定 利用者権限管理 アクセスコントロール アクセスコントロール機能 アクセス制御 アクセス制限 アクセス制御機構 アクセス制限情報 アクセス管理 ユーザアクセス制御 アクセス制御サーバ データアクセス制限	システム管理者権限 アクセス権限管理 権限管理 管理者の承認を要する設定 アクセス権限設定 アクセス権限の設定 アクセス権限の管理機能 アクセス権限の登録 アクセス権限を追加 ログイン試行回数の制限 アカウントの属性履歴 ログイン用アカウント アクセス数 エンドユーザのアカウント ユーザアカウント アカウントデータ 電子メールアカウント アプリケーションのアカウント ユーザ・アカウント管理 アカウント管理ツール ユーザ数 ユーザ制限機能システム アクセスログ データベースアクセスログ アクセスログ管理 アクセスログの管理 アクセスログを分析 アクセスログを監視 アクセスログ解析 アクセスログによる追跡機能 アクセスログを記録 アクセス履歴の管理 アクセス履歴等の管理

	アクセス権限	ファイルアクセス制御	アクセスログの記録
	操作権限	アクセス制御ソフトウェア	アクセス数
	権限レベル	アクセス管理基盤	アクセス履歴
	アクセス権限を管理	アクセスポリシー	アクセスログデータ
	入力等の権限	アクセス制御管理	アクセス解析データサーバ
	使用者権限設定	システム使用制限	
	操作者権限	アクセス権管理システム	
	権限を有するもの	アクセス権制御	
負荷バランス 語数:41	負荷分散	運用負荷	ネットワーク負荷分散
	負荷分散性能	負荷分散システム	回線負荷
	負荷分散処理	負荷分散方式	最大負荷時
	負荷分散機能	負荷分散構成	負荷率
	負荷分散環境	動的負荷分散機能	最小サーバ負荷
	負荷分散装置	負荷分散停止	ピーク時
	ロードバランス	サーバ負荷分散	ピーク時の負荷
	ロードバランシング	負荷分散クラスタ	高負荷
	ロード・バランシング	自律負荷バランス機能	負荷許容範囲
	ロードバランス設定	サーバロードバランサ	負荷監視
	ロードバランシング構成	CPU負荷	負荷検査
	負荷分散ポリシー	ネットワーク負荷	負荷予測
	負荷低減	ネットワークの負荷	負荷計測
	負荷運用	負荷分散対象ノード	
セキュリティ管理 語数:102	セキュリティ管理	インシデント管理ソフトウェア	セキュリティパッチ機能
	セキュリティ確保	インシデント管理機能	セキュリティパッチ提供
	セキュリティ運用	インシデント項目	セキュリティパッチ適合
	セキュリティの確保	インシデント発生分析	セキュリティパッチ配付
	セキュリティー確保	インシデント発生件数	セキュリティパッチ管理
	セキュリティの維持	インシデント番号	セキュリティ修正パッチ
	セキュリティ設定	インシデントレポート管理	セキュリティパッチ
	セキュリティ運用手順	インシデント管理業務フロー	セキュリティパッチ内容適合
	セキュリティ権限	インシデントレポート	セキュリティパッチ適用計画
	情報セキュリティ	インシデント管理	セキュリティパッチ対策ソフトウェア
	情報セキュリティ対策統一基準	インシデント発生状況	パッチあて
	情報セキュリティ対策基準	コンピュータセキュリティインシデント	修正パッチ
	情報セキュリティ実施手順書	監査証跡	パッチ適用
	情報セキュリティ実施手順	セキュリティ監査ツール	パッチプログラム
	セキュリティ実施手順	監査業務	パッチ評価
	情報セキュリティ管理	監査情報	データパッチ
	情報セキュリティマネジメントシステム	セキュリティ監査	手動パッチ
	情報セキュリティ規則	情報セキュリティ監査報告書	パッチツール
	情報セキュリティ対策要領	セキュリティ監査ログ	パッチ対応
	情報セキュリティ運用マニュアル	情報セキュリティ監査	パッチファイル
	セキュリティ管理システム	情報セキュリティポリシー	バグ修正パッチ
	セキュリティ管理機能	セキュリティポリシー	OSパッチ処理
	セキュリティ管理計画書	検疫ポリシー変更手順	プログラム修正パッチ
	セキュリティ管理要領	ローカルセキュリティポリシー	システムパッチ
	セキュリティガイドライン	セキュリティ方針書	システムパッチログ
	情報セキュリティ管理区域	情報セキュリティ教育	パッチ管理
	セキュリティレベル	情報セキュリティ研修	パッチ情報
	セキュリティレベル	セキュリティ教育	プログラム修正パッチ適用対応
	セキュリティ管理レベル	セキュリティに関する研修	パッチ適用管理機能
	セキュリティ対策レベル	セキュリティリスク	パッチファイル適用状況
	情報セキュリティレベル	セキュリティイベント件数	システムパッチ実施状況
	インシデント管理システム	セキュリティ対策適用箇所一覧	修正パッチ適用環境
	インシデント管理サーバ	セキュリティチェック項目	オンラインパッチ処理
	最新セキュリティパッチ	セキュリティリスク分析機能	
障害対策 語数:76	障害処理	ホットプラグ機能	障害管理
	障害対策	電源管理ソフトウェア	障害管理機能
	システム障害対策	障害未然防止	障害管理プロセス
	障害対応方法	障害予防	耐障害性
	障害対応	障害防止管理	障害事後対策
	障害連絡	障害再発防止	障害判定基準
	故障対策	落雷対策	障害管理サーバ
	故障対応	雷対策	障害発生日時
	エラー処理	停電装置	障害発生頻度
	マルウェア対策	停電対策	障害再発限度
	マルウェア対策ソフトウェア	障害発生機器	障害発生限度
	障害調査	障害再発	障害分類毎障害件数

	システム障害対応	障害再発防止	障害対応工数
	トラブルシューティング	障害検知ソフトウェア	ハードウェア障害対応マニュアル
	365日障害対応	障害履歴	障害対応マニュアル
	緊急対応機能	障害票	障害手順書
	障害対応状況報告書	障害対応結果報告書	情報システムダウン対策マニュアル
	故障対応状況	障害管理台帳	障害復旧手順
	計画停電	アクシデントレポート作成機能	回線障害対応手順
	計画停電対応	障害報告書	サーバ障害対応手順
	計画停止スケジュール	障害記録	不正侵入障害対応手順
	計画停電テスト	障害管理簿	障害対策マニュアル
	フェイルオーバーテスト	エラーログ管理	トラブルシューティング機能
	フェイルオーバー機能	障害管理計画	障害切り分けマニュアル
	リンクアグリゲーション	障害対応計画	
	ホットスワップ	障害管理方法	